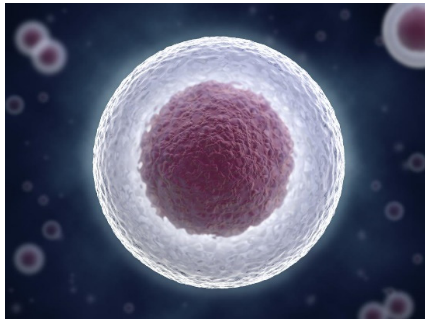
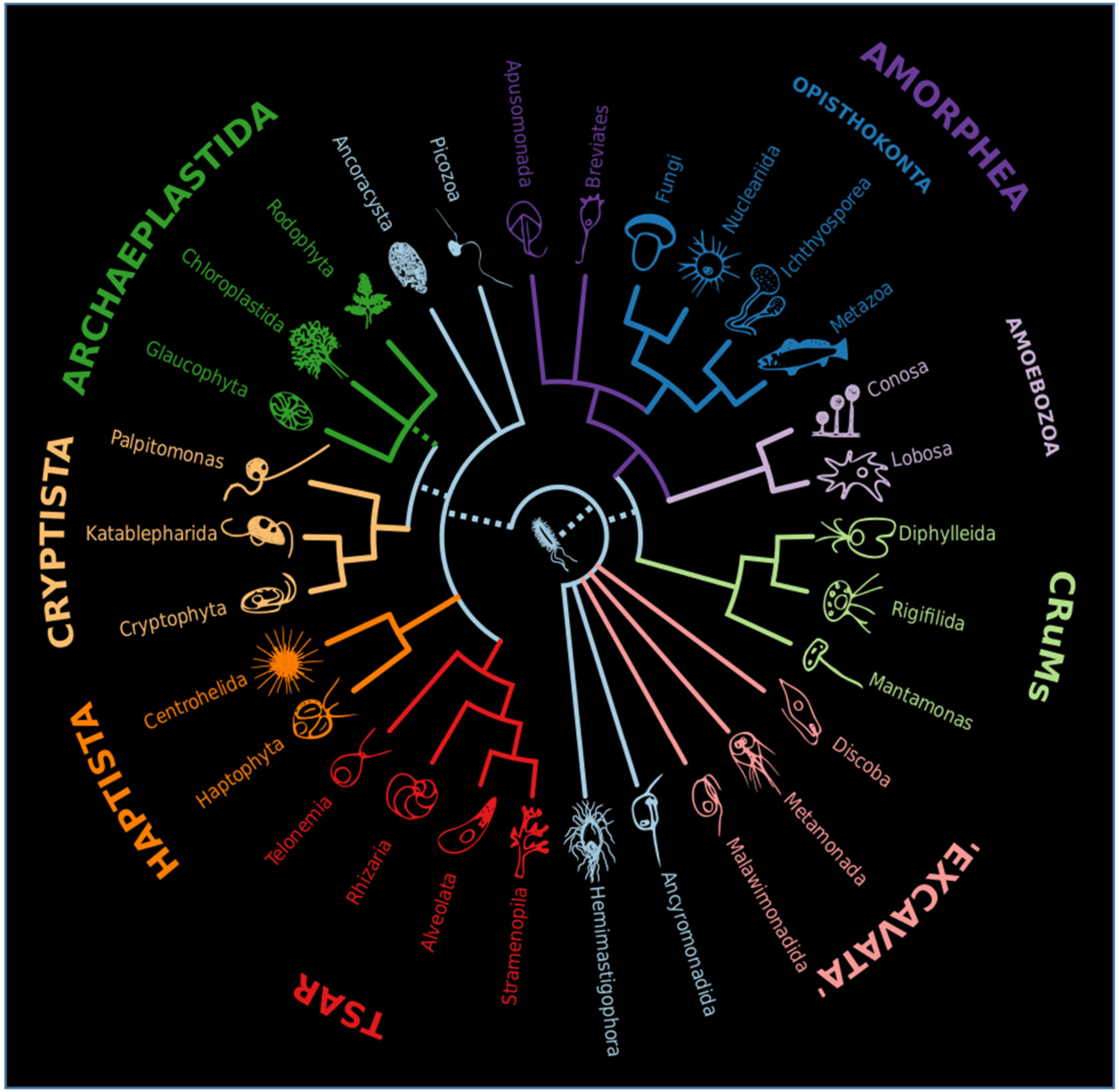
Genome and chromosome evolution



Martin A. Lysák

CEITEC, Masaryk University

www.plantcytogenomics.org

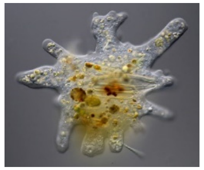

ARCHAEPLASTIDA
Rodophyta
Chloroplastida
Glaucophyta
Ancoracysta
Picozoa
Apusomonada
Breviates
Fungi
Nucleariida
Ichthyosporea
Metazoa
OPISTHOKONTA
AMORPHEA
Conosa
Lobosa
AMOEBOZOA
Diphylleida
Rigifilida
Mantamonas
CRuMS
Discoba
Metamonada
Malawimonadida
'EXCAVATA'
Ancyromonadida
Hemimastigophora
Stramenopila
Alveolata
Rhizaria
Telonemia
TSAR
Haptophyta
Centrohelida
HAPTISTA
Cryptophyta
Katablepharida
Palpitomonas
CRYPTISTA

Genome and chromosome evolution

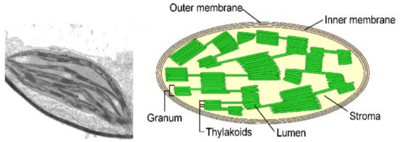
- genome size variation
- chromosome number variation
- variation in coding DNA amount
- variation in non-coding DNA

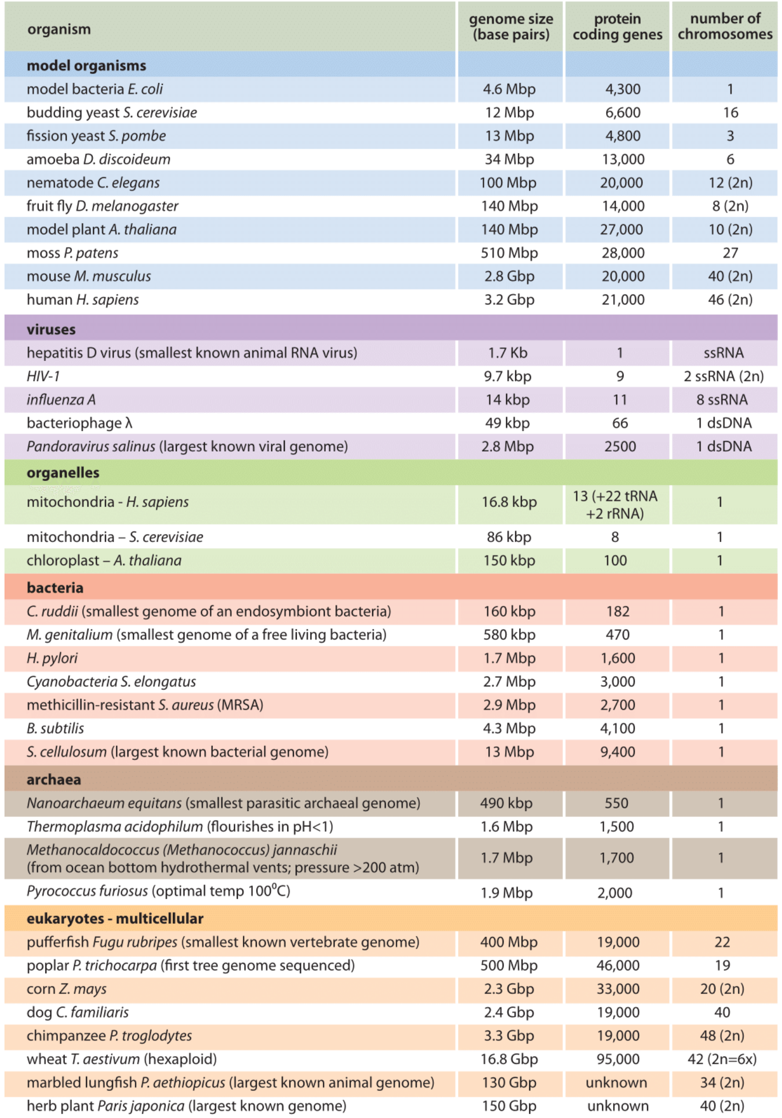
Genome size variation

Polychaos dubium

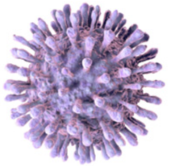


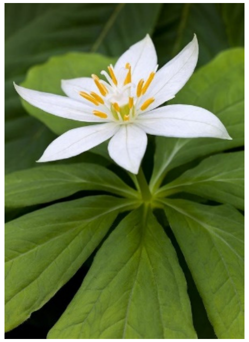
…perhaps the largest known genome –
670 billion base pairs (670 Gb)
(~200-times larger than the human genome, 3.2 Gb; some authors suggest treating the value with caution)





organism	genome size (base pairs)	protein coding genes	number of chromosomes
model organisms			
model bacteria *E. coli*	4.6 Mbp	4,300	1
budding yeast *S. cerevisiae*	12 Mbp	6,600	16
fission yeast *S. pombe*	13 Mbp	4,800	3
amoeba *D. discoideum*	34 Mbp	13,000	6
nematode *C. elegans*	100 Mbp	20,000	12 (2n)
fruit fly *D. melanogaster*	140 Mbp	14,000	8 (2n)
model plant *A. thaliana*	140 Mbp	27,000	10 (2n)
moss *P. patens*	510 Mbp	28,000	27
mouse *M. musculus*	2.8 Gbp	20,000	40 (2n)
human *H. sapiens*	3.2 Gbp	21,000	46 (2n)
viruses			
hepatitis D virus (smallest known animal RNA virus)	1.7 Kb	1	ssRNA
HIV-1	9.7 kbp	9	2 ssRNA (2n)
influenza A	14 kbp	11	8 ssRNA
bacteriophage λ	49 kbp	66	1 dsDNA
Pandoravirus salinus (largest known viral genome)	2.8 Mbp	2500	1 dsDNA
organelles			
mitochondria - *H. sapiens*	16.8 kbp	13 (+22 tRNA +2 rRNA)	1
mitochondria – *S. cerevisiae*	86 kbp	8	1
chloroplast – *A. thaliana*	150 kbp	100	1
bacteria			
C. ruddii (smallest genome of an endosymbiont bacteria)	160 kbp	182	1
M. genitalium (smallest genome of a free living bacteria)	580 kbp	470	1
H. pylori	1.7 Mbp	1,600	1
Cyanobacteria S. elongatus	2.7 Mbp	3,000	1
methicillin-resistant *S. aureus* (MRSA)	2.9 Mbp	2,700	1
B. subtilis	4.3 Mbp	4,100	1
S. cellulosum (largest known bacterial genome)	13 Mbp	9,400	1
archaea			
Nanoarchaeum equitans (smallest parasitic archaeal genome)	490 kbp	550	1
Thermoplasma acidophilum (flourishes in pH<1)	1.6 Mbp	1,500	1
Methanocaldococcus (Methanococcus) jannaschii (from ocean bottom hydrothermal vents; pressure >200 atm)	1.7 Mbp	1,700	1
Pyrococcus furiosus (optimal temp 100°C)	1.9 Mbp	2,000	1
eukaryotes - multicellular			
pufferfish *Fugu rubripes* (smallest known vertebrate genome)	400 Mbp	19,000	22
poplar *P. trichocarpa* (first tree genome sequenced)	500 Mbp	46,000	19
corn *Z. mays*	2.3 Gbp	33,000	20 (2n)
dog *C. familiaris*	2.4 Gbp	19,000	40
chimpanzee *P. troglodytes*	3.3 Gbp	19,000	48 (2n)
wheat *T. aestivum* (hexaploid)	16.8 Gbp	95,000	42 (2n=6x)
marbled lungfish *P. aethiopicus* (largest known animal genome)	130 Gbp	unknown	34 (2n)
herb plant *Paris japonica* (largest known genome)	150 Gbp	unknown	40 (2n)



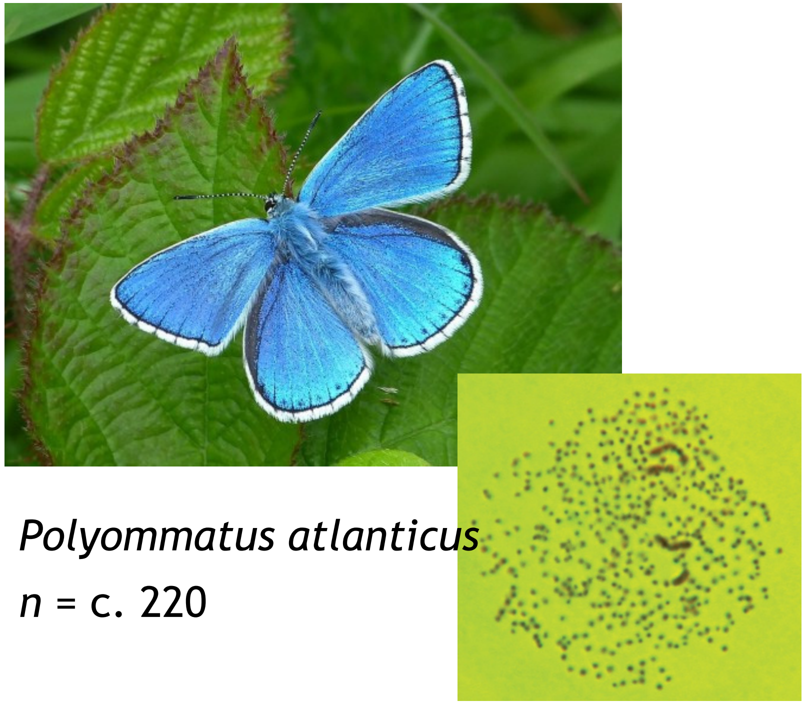


Eukaryotes: minimal chromosome numbers

Myrmecia pilosula „Jack jumper ant“,
Australia; males (haploid) $n = 1$, females
(diploid) $2n = 2$

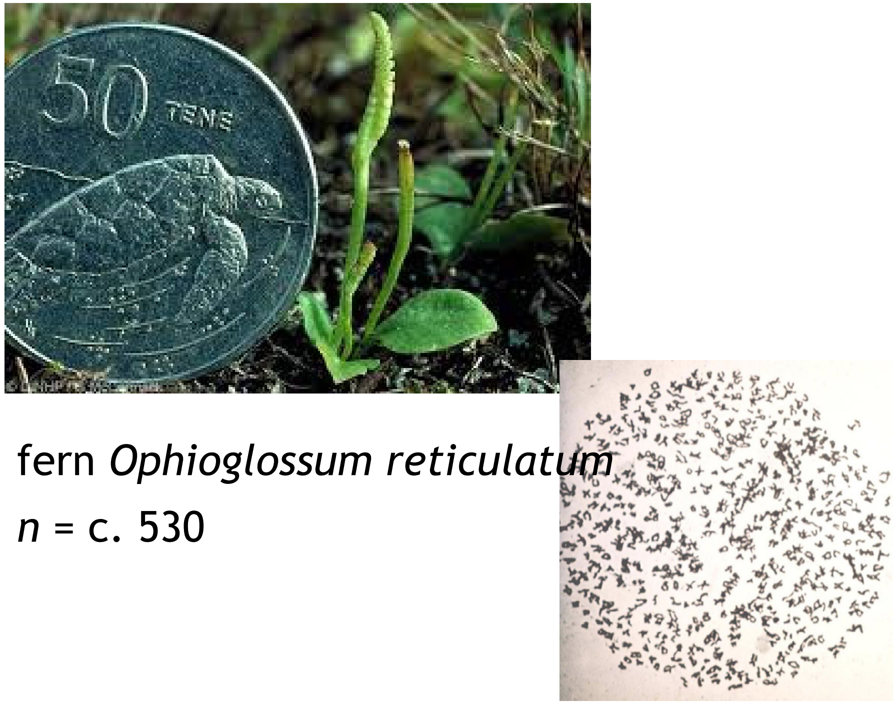
five angiosperm species
e.g., *Haplopappus gracilis*, Asteraceae,
$n = 2$

Eukaryotes: highest chromosome numbers



Polyommatus atlanticus

n = c. 220



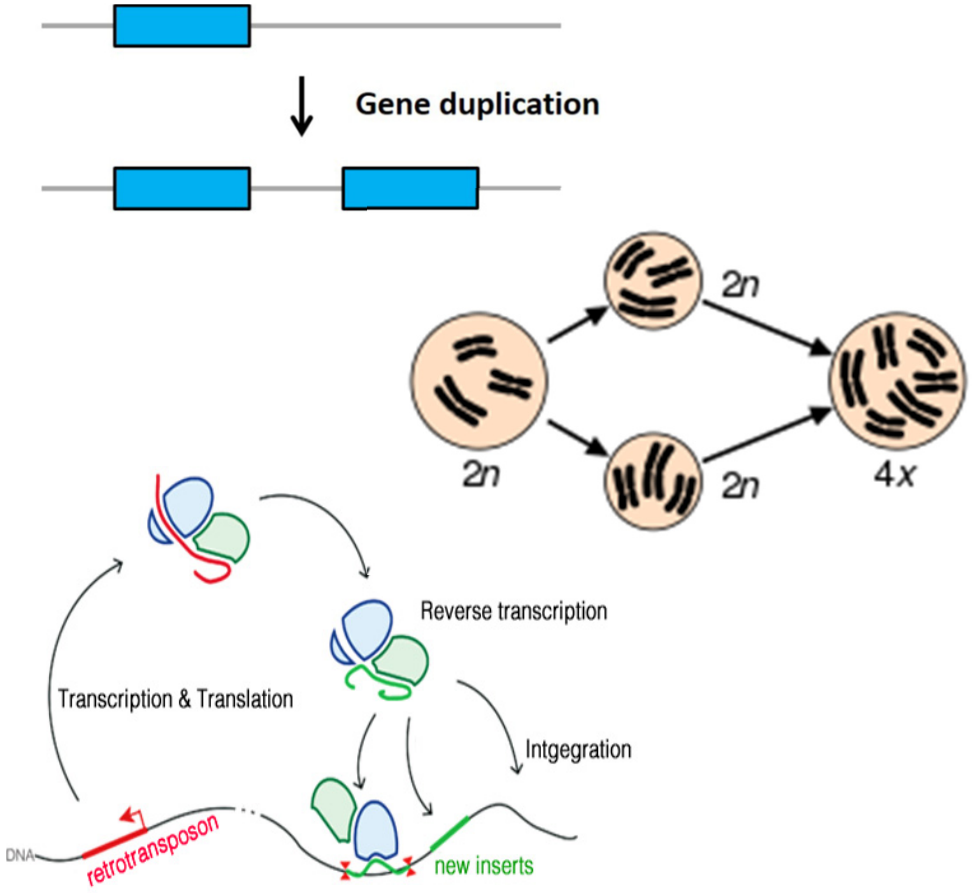
fern *Ophioglossum reticulatum*

n = c. 530

Genome size ⟺ Chromosome number

Variation in genome size and chromosome number is driven by two principal processes

DNA/genome duplication



DNA recombination

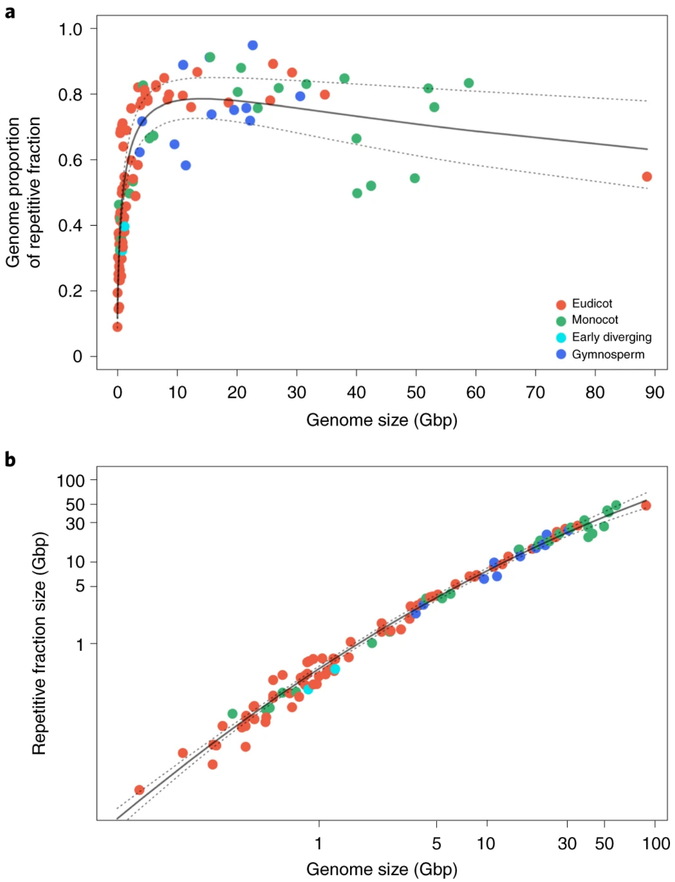
⇩

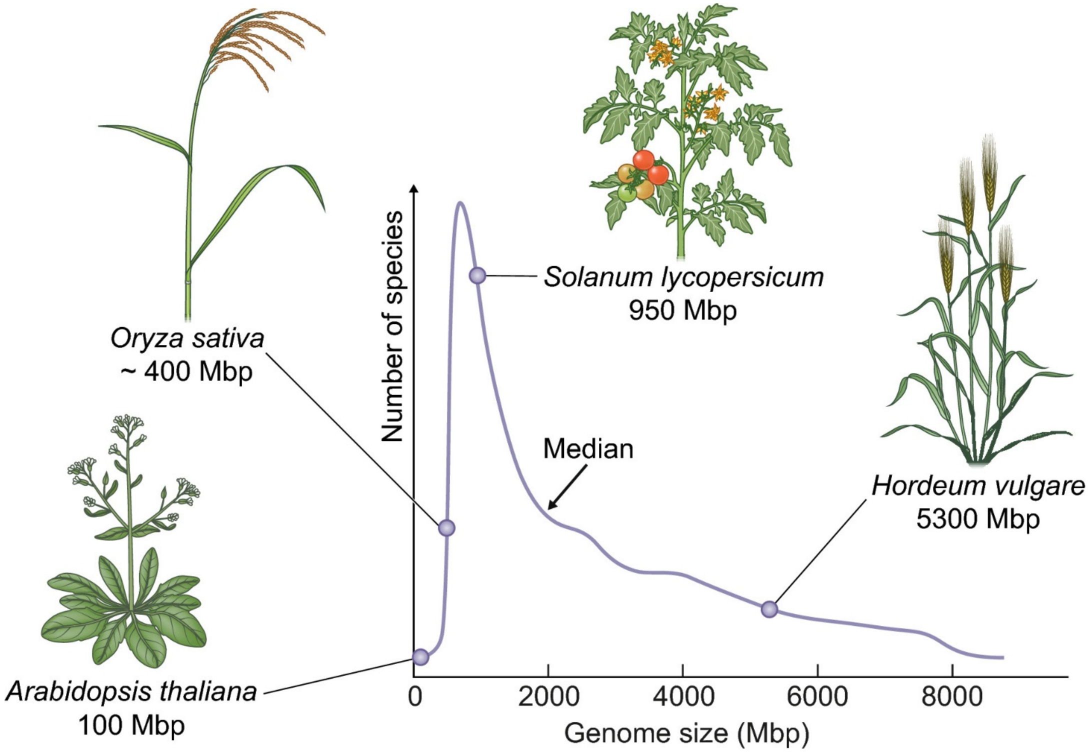
recombination

Genome size increase

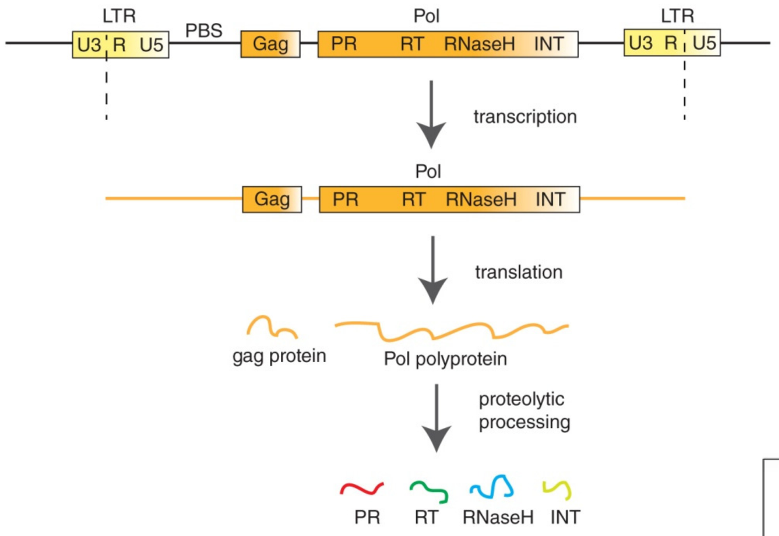
- amplification of retrotransposons (and tandem repeats)
- gene and segmental duplications
- polyploidy

Genome size variation in angiosperms is driven by amplification (and elimination) of repetitive DNA





LTR (Long Terminal Repeat) retrotransposons (LTR-RTs)



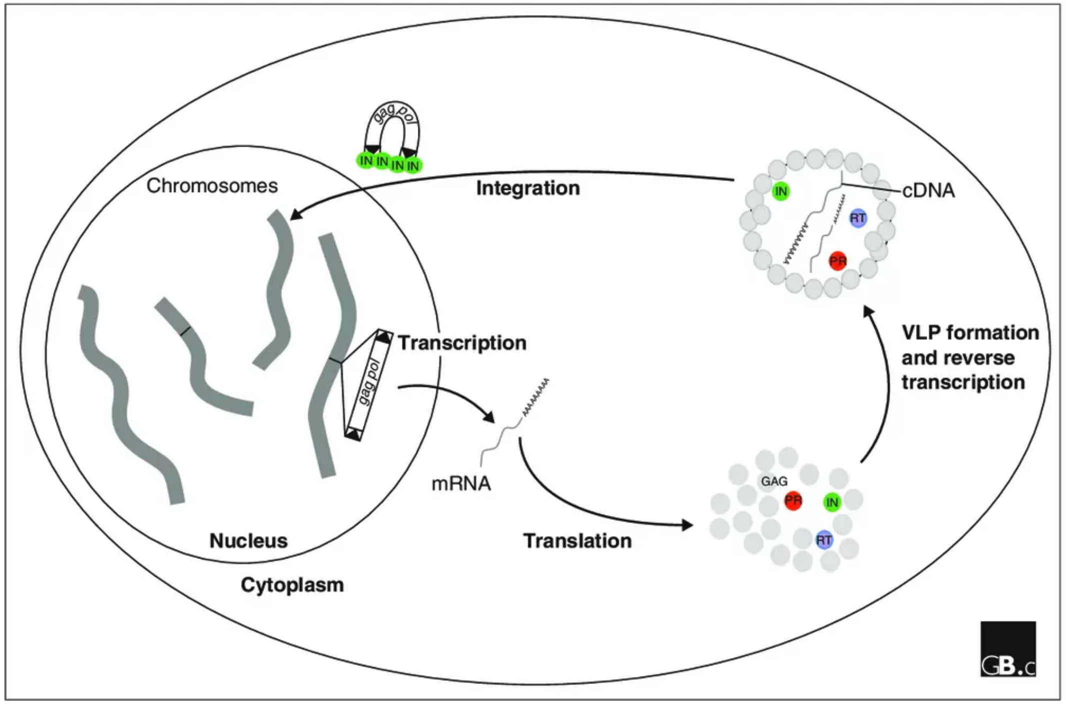
Gag - gene for the Gag protein

INT - integrase

PBS - primer binding site

PR - protease

RT - reverse transcriptase



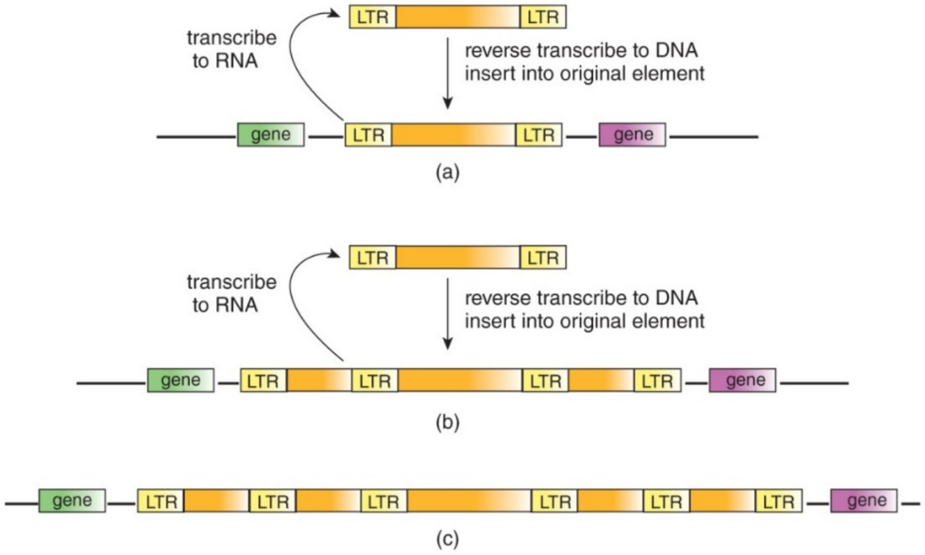
VLP - virus-like particle

Genome size increase by retrotransposition (nested retrotransposon insertion)

(a) rt is inserted into itself

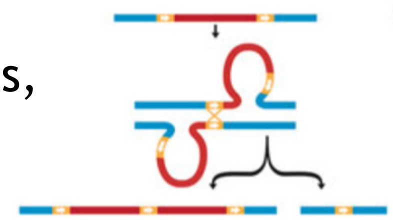
(b) the event is repeated

(c) DNA structure after 2 rounds of retrotransposition

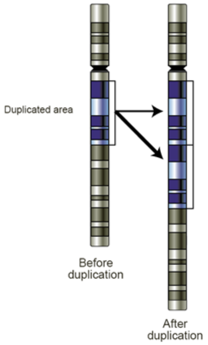


Genome size increase by gene duplication

- replication slippage (errors in replication → gene duplication)



- ectopic recombination (between two direct repeats, typically TEs)



- unequal crossing-over in meiosis (due to missaligned chromosomes)
- via retrotransposition = retrogenes (cellular mRNA is transcribed into cDNA by reverse transcriptase of a retrotransposon or retrovirus; retrogene does not contain introns = lacking regulatory elements = pseudogene, but can evolve into a functional gene)

Retrogenes

- mRNA is reverse-transcribed into cDNA and inserted in a new genomic position

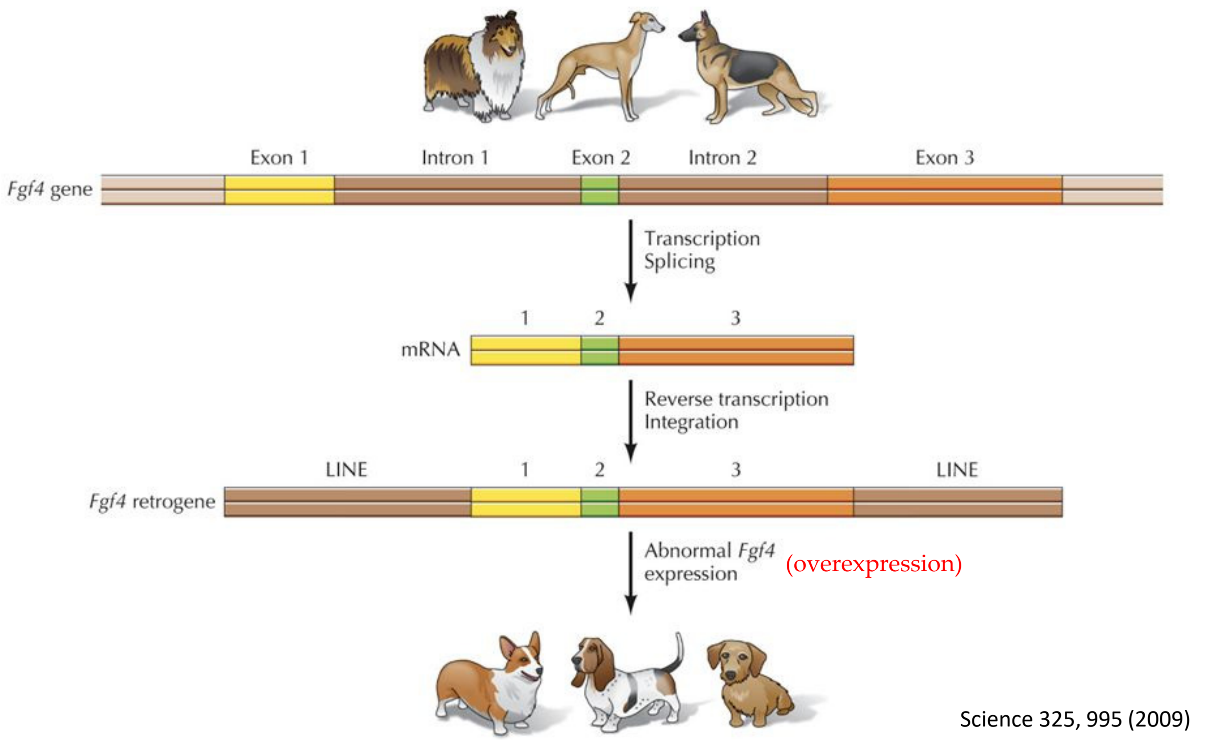
inactive gene copy

retrogene

chimeric retrogene

(functional by capturing elements of another gene)

Figure 6.15 Transposition of a retrogene determines short legs in dog breeds



Science 325, 995 (2009)

Segmental duplications

- duplicated segment of chromosomal DNA (usually defined as > 1 kb in length)
- either tandem or interspersed organization, either intra-chromosomal or inter-chromosomal
- also known as low copy repeats
- human genome: 159 Mb gene-rich duplicated (5.5% of the genome) = c. Arabidopsis genome

Variation in the segmentally duplicated amylase locus in humans

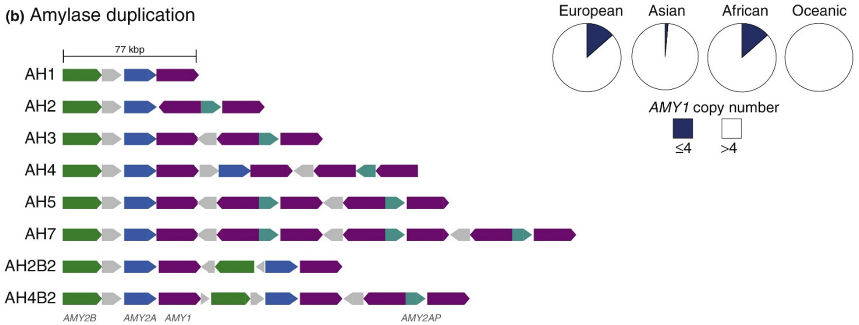
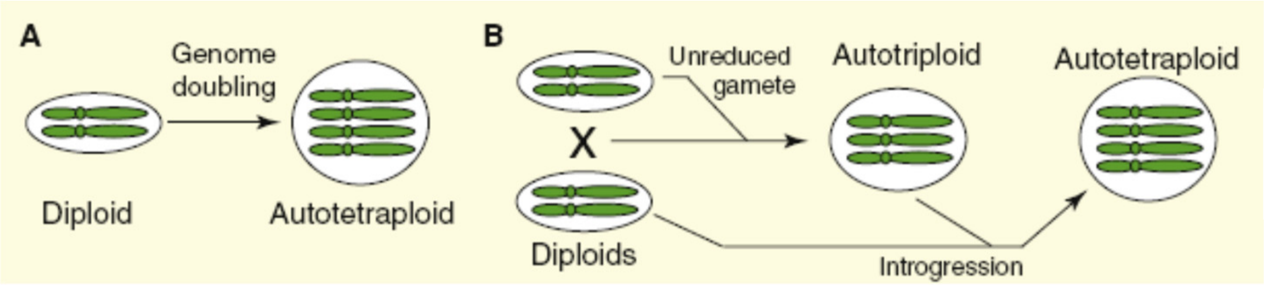

(b) Amylase duplication

Table 1 | **SD content of sequenced animal genomes**

	Caenorhabditis elegans	***Drosophila melanogaster***	**Human**	**Mouse**	**Rat**	**Chicken**	**Chimpanzee***
SDs of >1 kb	4.3%	1.2%	5.2%	2.7%	1.6%	2.7%	N.D.
SDs of >10 kb	0.7%	0.1%	4.5%	2.2%	1.5%	0.3%	N.D.
SDs of >20 kb	N.D.	N.D.	4.0%	1.7%	0.9%	0.0%	~4.8%
Genome size	97	123	2,866	2,506	2,566	1,040	2,866

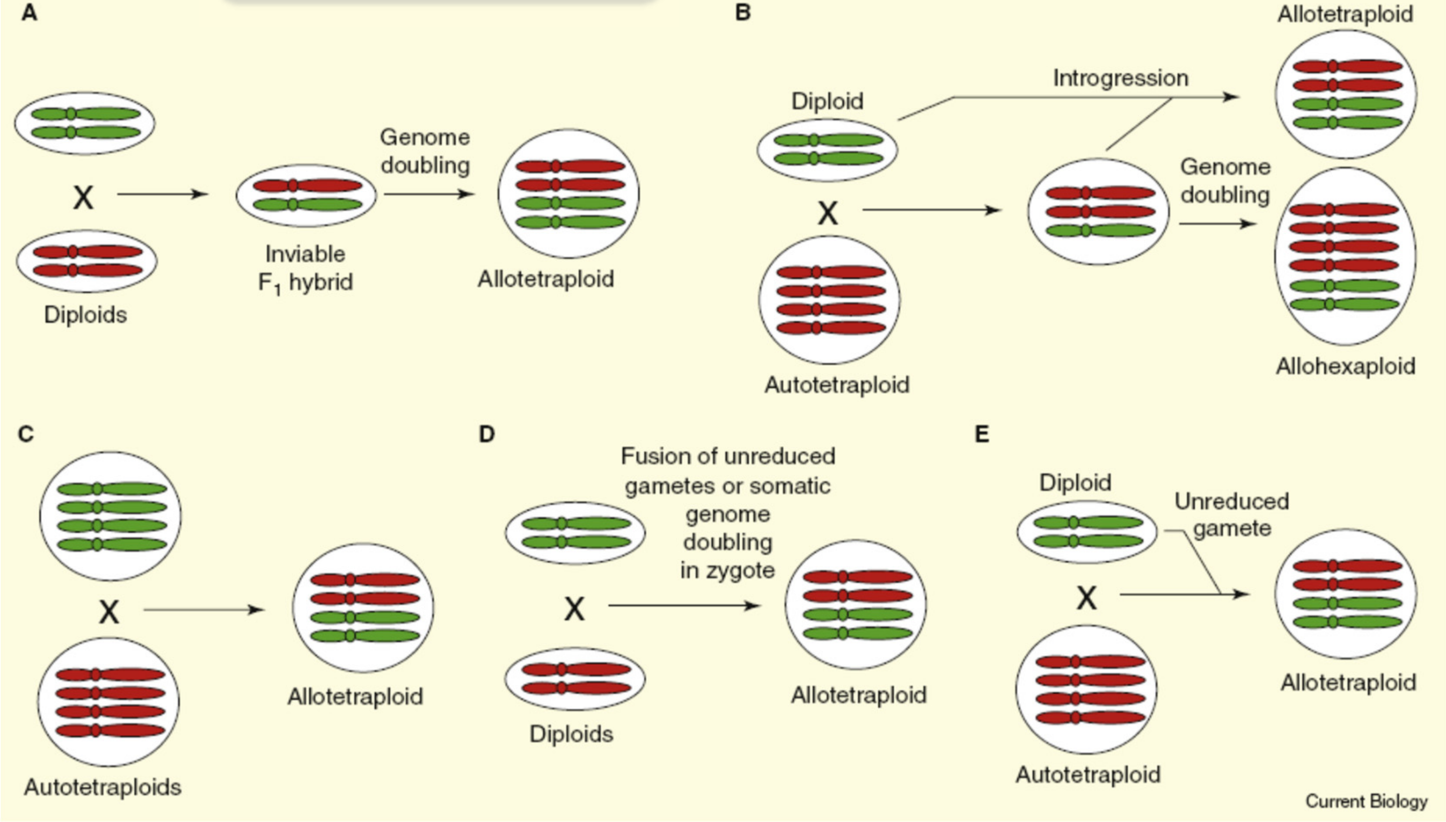
Data taken from REFS 2,7 for pairwise segmental duplications (SDs) with >90% identity. *Given the fragmented nature of SDs in the draft chimpanzee genome, the duplication content can only be estimated indirectly on the basis of human duplication content, adjusting for detected differences in SD compared with chimpanzee whole-genome shotgun sequencing[6]. DNA not assigned to a chromosome was not included in these calculations. Consequently, in other genomes the estimate of recent duplication might rise as the quality of the sequence assembly improves. N.D., not determined.

Polyploidy (whole-genome duplication)



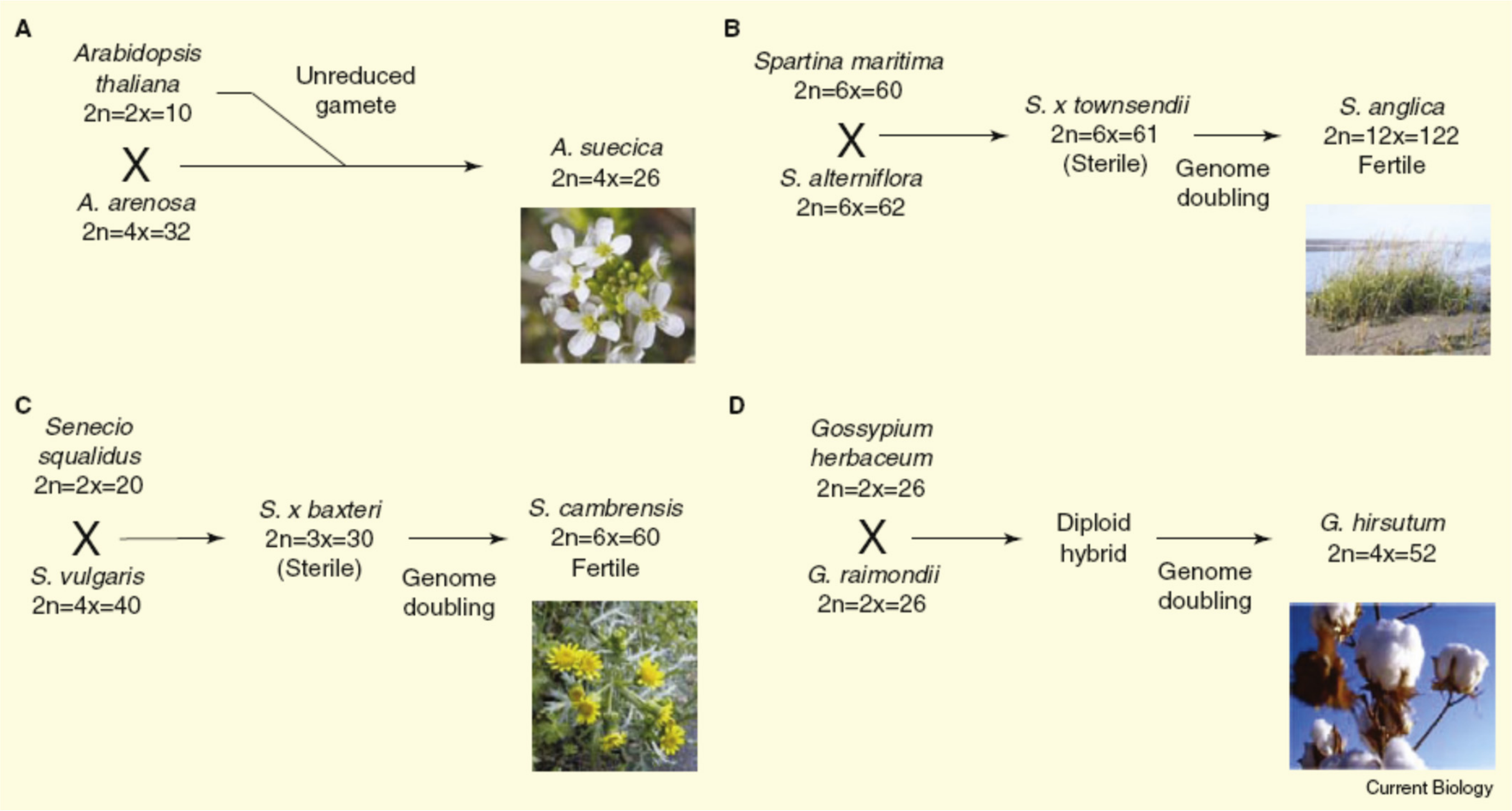
AUTOPOLYPLOIDY

ALLOPOLYPLOIDY



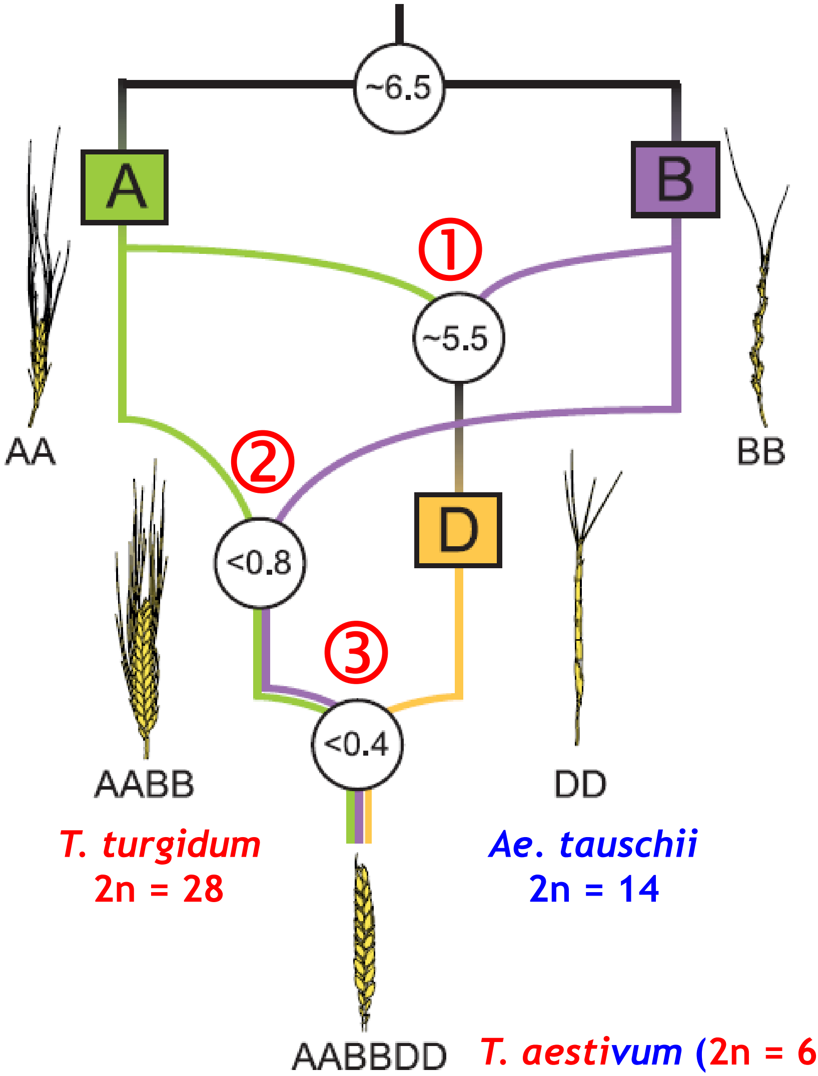
Hegarty and Hiscock 2008, Current Biol

Examples of allopolyploid speciation



Hegarty and Hiscock 2008,
Current Biol

Phylogenomic history of bread wheat (*Triticum aestivum*; AABBDD).
Three rounds of hybridization/polyploidy.



T. turgidum
2n = 28

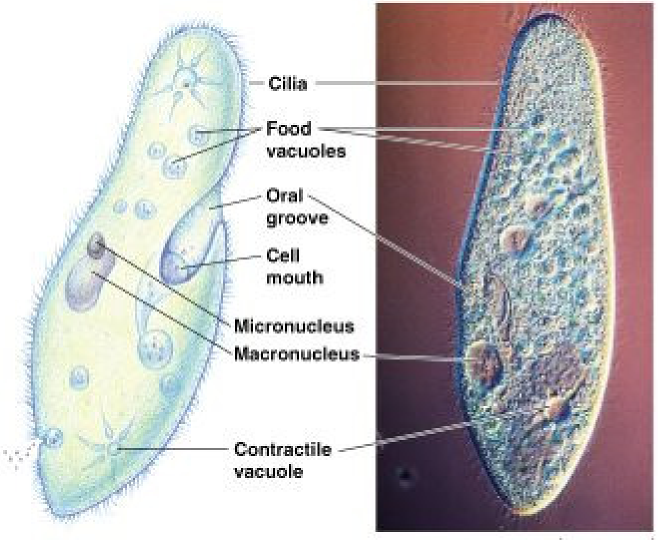
Ae. tauschii
2n = 14

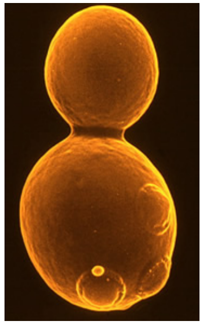
T. aestivum (2n = 6x = 42)

Marcussen et al. (2014), Science

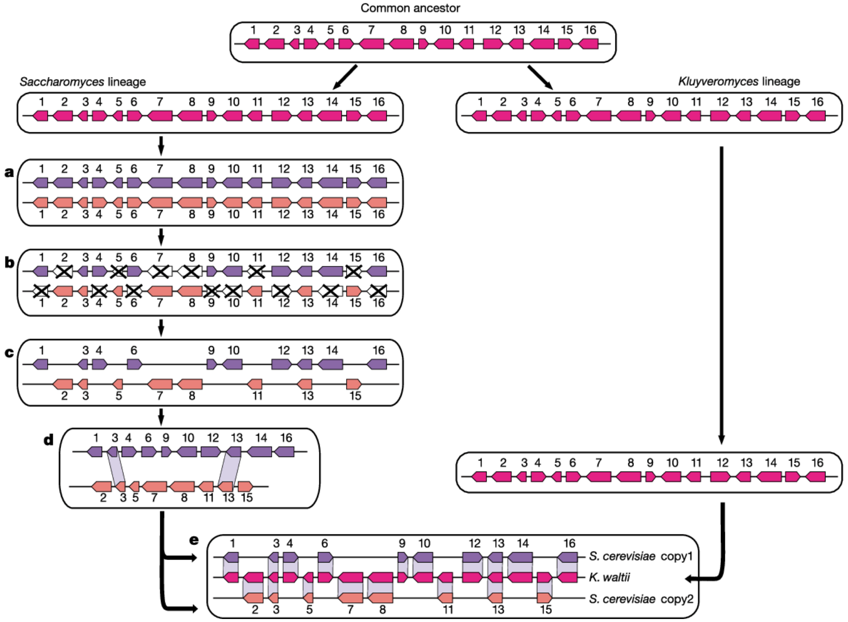
Whole-genome duplications in protozoa

- the unicellular eukaryote *Paramecium tetraurelia*
- most of 40,000 genes arose through at least 3 successive whole-genome duplications
- most recent duplication most likely caused an explosion of speciation events that gave rise to the *P. aurelia* complex (15 sibling species)
- some genes have been lost, some retained
- many retained (duplicated) genes do not generate functional innovations but are important because of the **gene dosage effect**





Whole-genome duplications in yeast



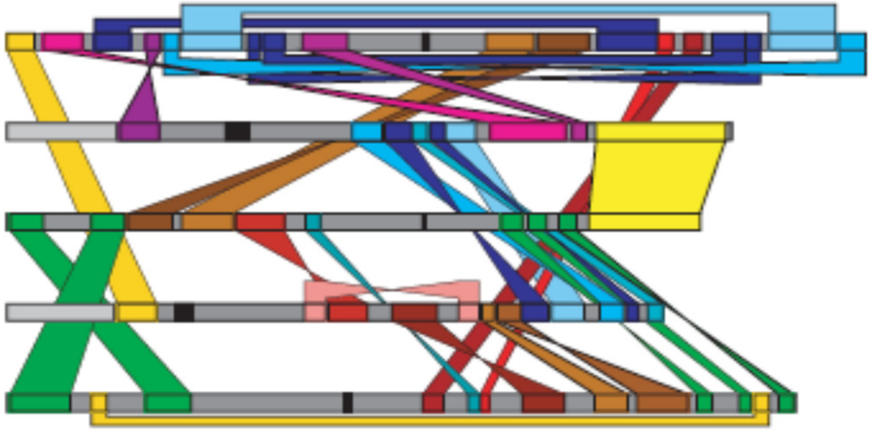
- genome comparison between two yeast species, *Saccharomyces cerevisiae* (n = 16) and *Kluyveromyces waltii* (n = 8)
- each region of *K. waltii* corresponding to two regions of *S. cerevisiae*
- **the *S. cerevisiae* genome underwent a WGD after the two yeast species diverged**
- in nearly every case (95%), accelerated evolution was confined to only one of the two paralogues (= one of the paralogues retained an ancestral function, the other was free to evolve more rapidly and acquired a derived function)

Kellis *et al*. 2004, Nature

First evidence of a WGD in plants

Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*

The Arabidopsis Genome Initiative* AGI (2000)

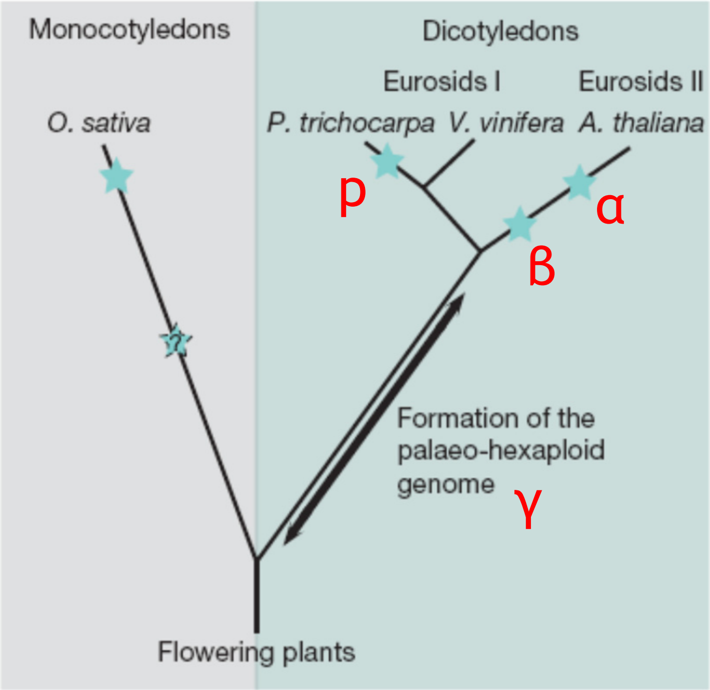


What does the duplication in the Arabidopsis genome tell us about the ancestry of the species? As the majority of the Arabidopsis genome is represented in duplicated (but not triplicated) segments, it appears most likely that **Arabidopsis, like maize, had a tetraploid ancestor** …The diploid genetics of Arabidopsis and the extensive divergence of the duplicated segments have masked its evolutionary history.

The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla

The French–Italian Public Consortium for Grapevine Genome Characterization*

Nature 449, 2007

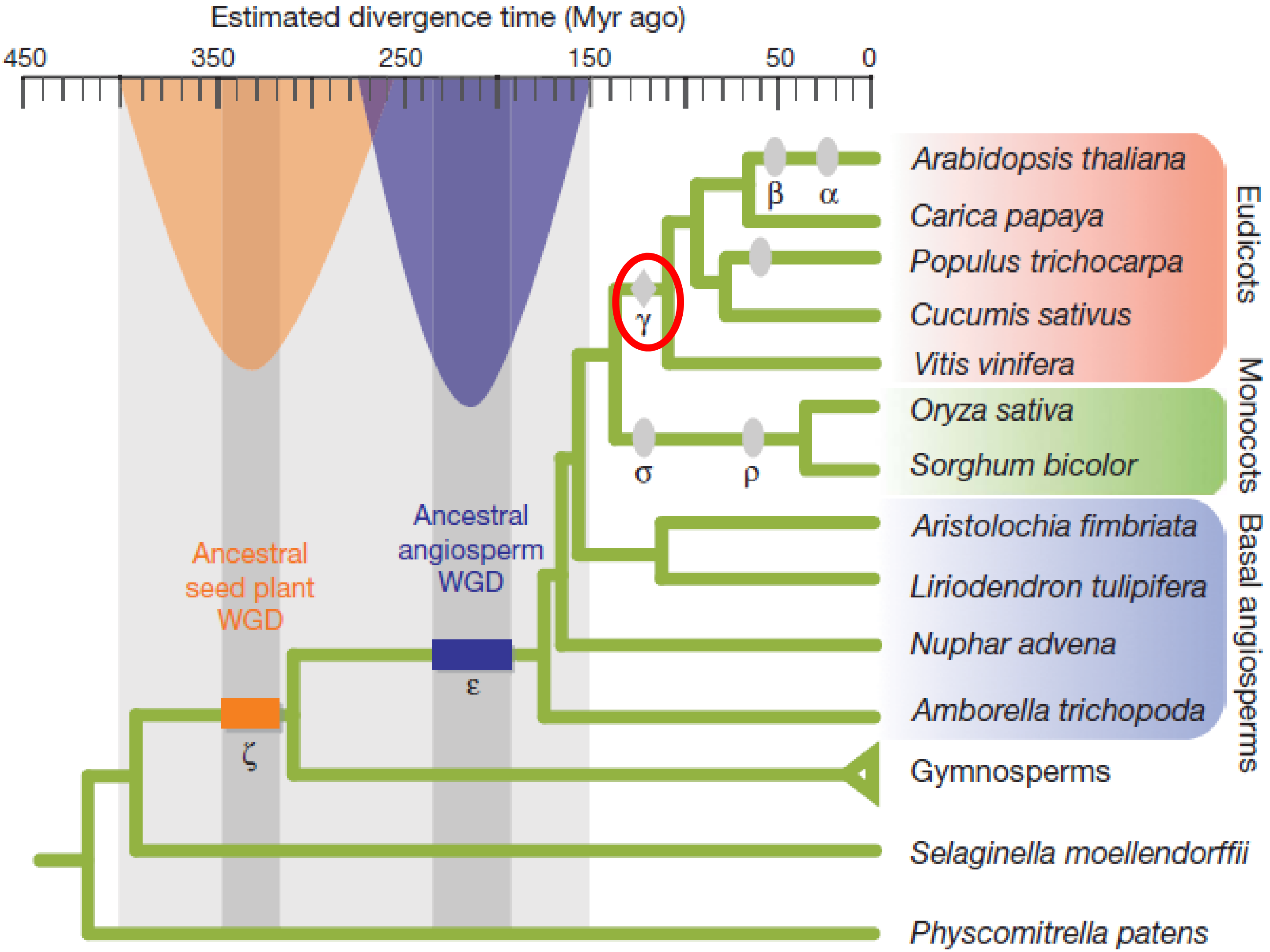


The formation of the palaeo-hexaploid ancestral genome occurred after divergence from monocots and before the radiation of the Eurosids. Star = a WGD (tetraploidization) event.

The **γ triplication** may have been an ancient **auto-hexaploidy** formed from fusions of three identical genomes, or **allo-hexaploidy** formed from fusions of three somewhat diverged genomes.

Tang *et al.* 2008, Genome Res

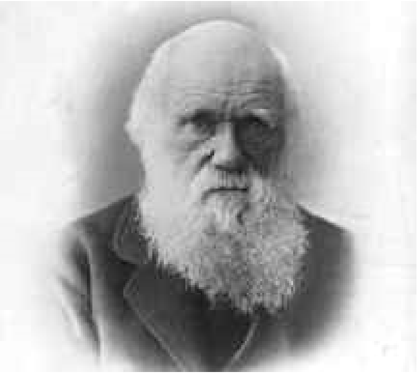
WGD events in seed plants and angiosperms



399 - 381 **319 - 297 Myr ago**

Jiao et al. (2011) Nature; Clark and Donoghue (2017) Proc R Soc

Charles Darwin’s abominable mystery solved (?)

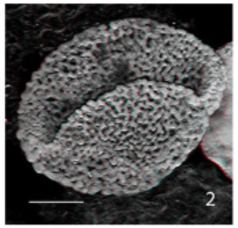


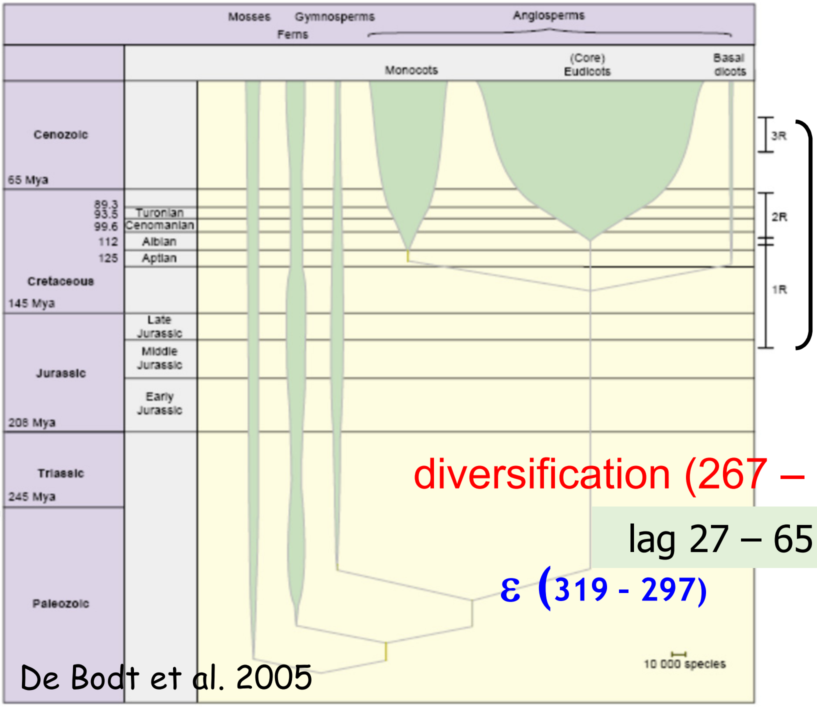
"The rapid development as far as we can judge of all the higher plants within recent geological times is an abominable mystery."

(Charles Darwin in a letter to Sir Joseph Hooker, 1879)

Archaefructus liaoningensis
(140 million year old fossil)

Afropollis
(245 million year old angiosperm pollen)





De Bodt et al. 2005

assumed ancient whole-genome duplication events
(e.g. γ - gamma WGD)

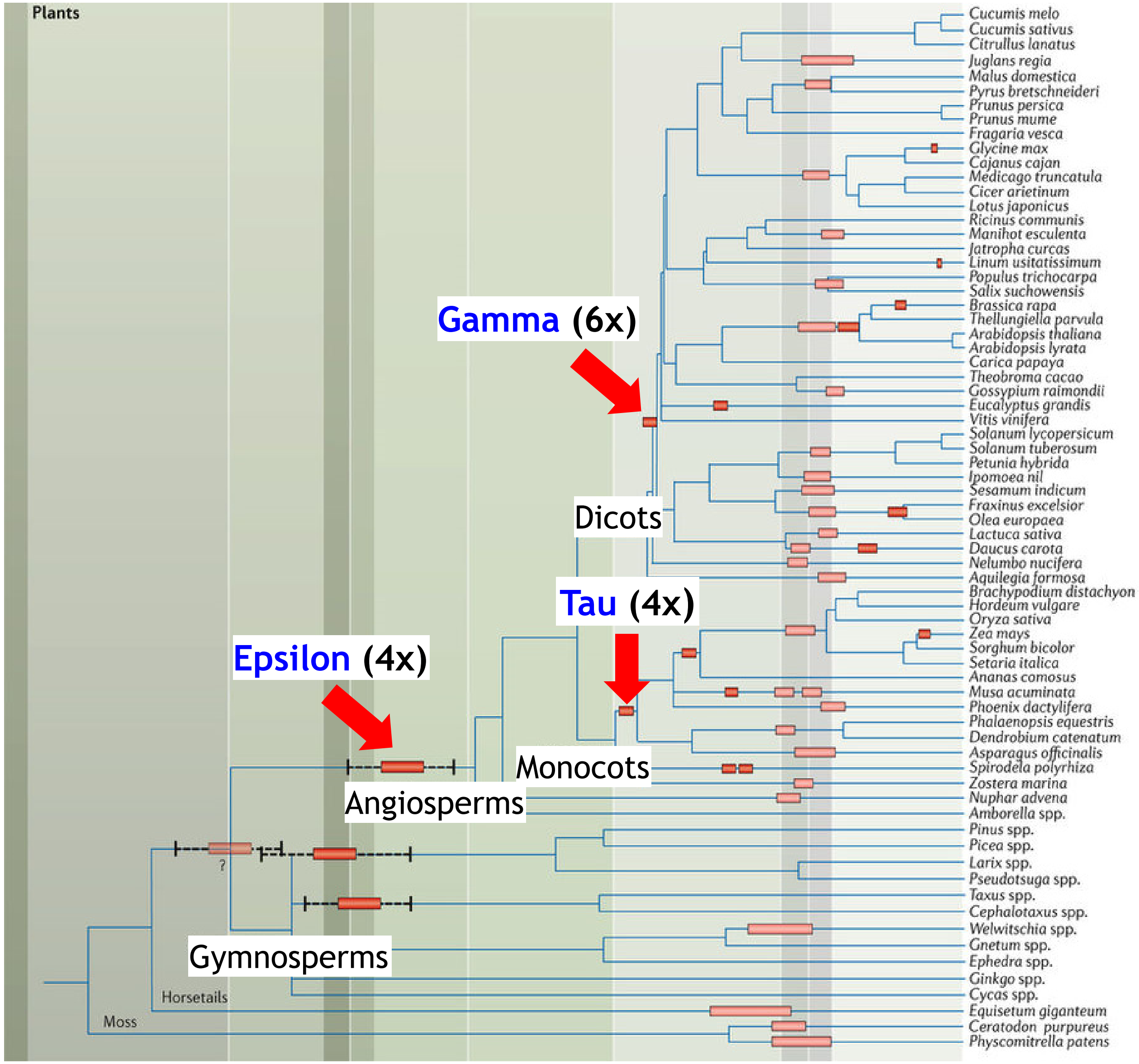
diversification (267 – 247)

lag 27 – 65 million years

ε (319 - 297)

Theres is evidence of ancient polyploidy throughout the major angiosperm lineages. It means that a genome-scale duplication event probably occurred PRIOR to the rapid diversification of flowering plants

Multiple whole-genome duplications in evolution of land plants



Van de Peer (2017) Nat Rev Genet (modified)

Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event

Jeffrey A. Fawcett[a,b,1], Steven Maere[a,b,1], and Yves Van de Peer[a,b,2]

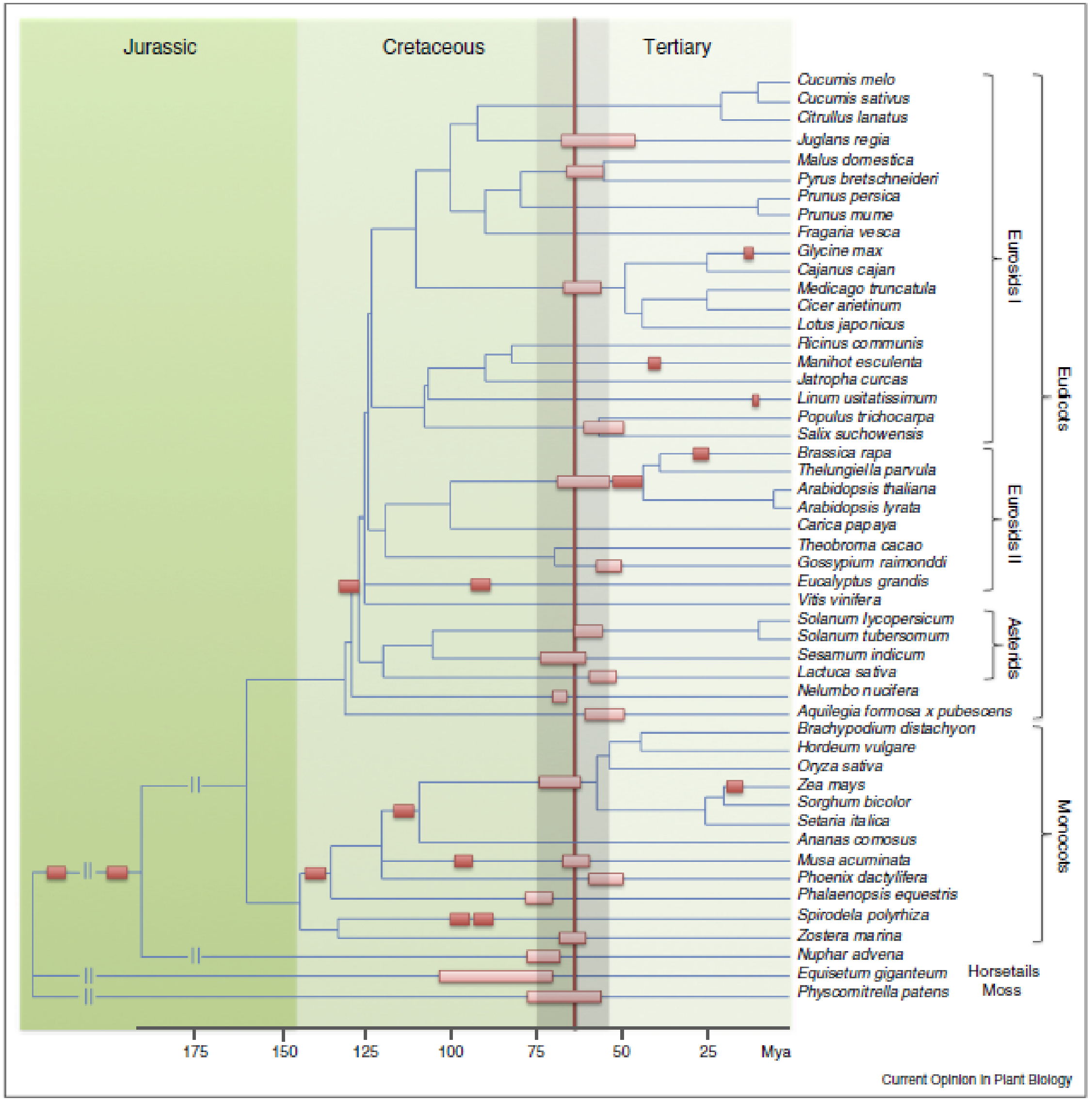
[a]Department of Plant Systems Biology, Flanders Institute for Biotechnology, 9052 Gent, Belgium; and [b]Department of Plant Biotechnology and Genetics, Ghent University, 9052 Gent, Belgium

PNAS 106 (2009)

Could WGD event(s) help plants to survive the mass extinction (one or more catastrophic events such as a massive asteroid impact) at the Cretaceous-Tertiary boundary ?

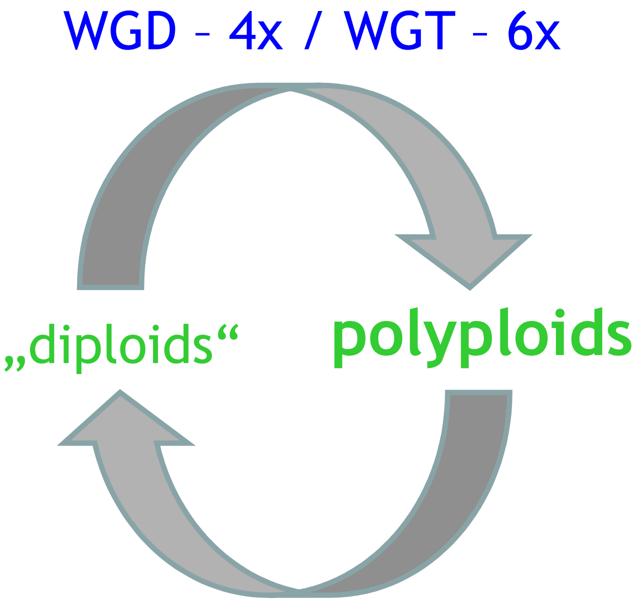
Possible establishment of polyploid plants following the K/Pg mass extinction (66 million y. ago)

- ➢ WGDs clustered around the Cretaceous–Tertiary (KT) boundary
- ➢ the KT extinction event - the most recent mass extinction (one or more catastrophic events such as a massive asteroid impact and/or increased volcanic activity)
- ➢ the KT extinction event - extinction of 60% of plant species, as well as a majority of animals, including dinosaurs



Lohaus and Van de Peer (2016) Curr Opin Pl Biol

Polyploidization – Diploidization cycle



WGD – 4x / WGT – 6x

„diploids“

polyploids

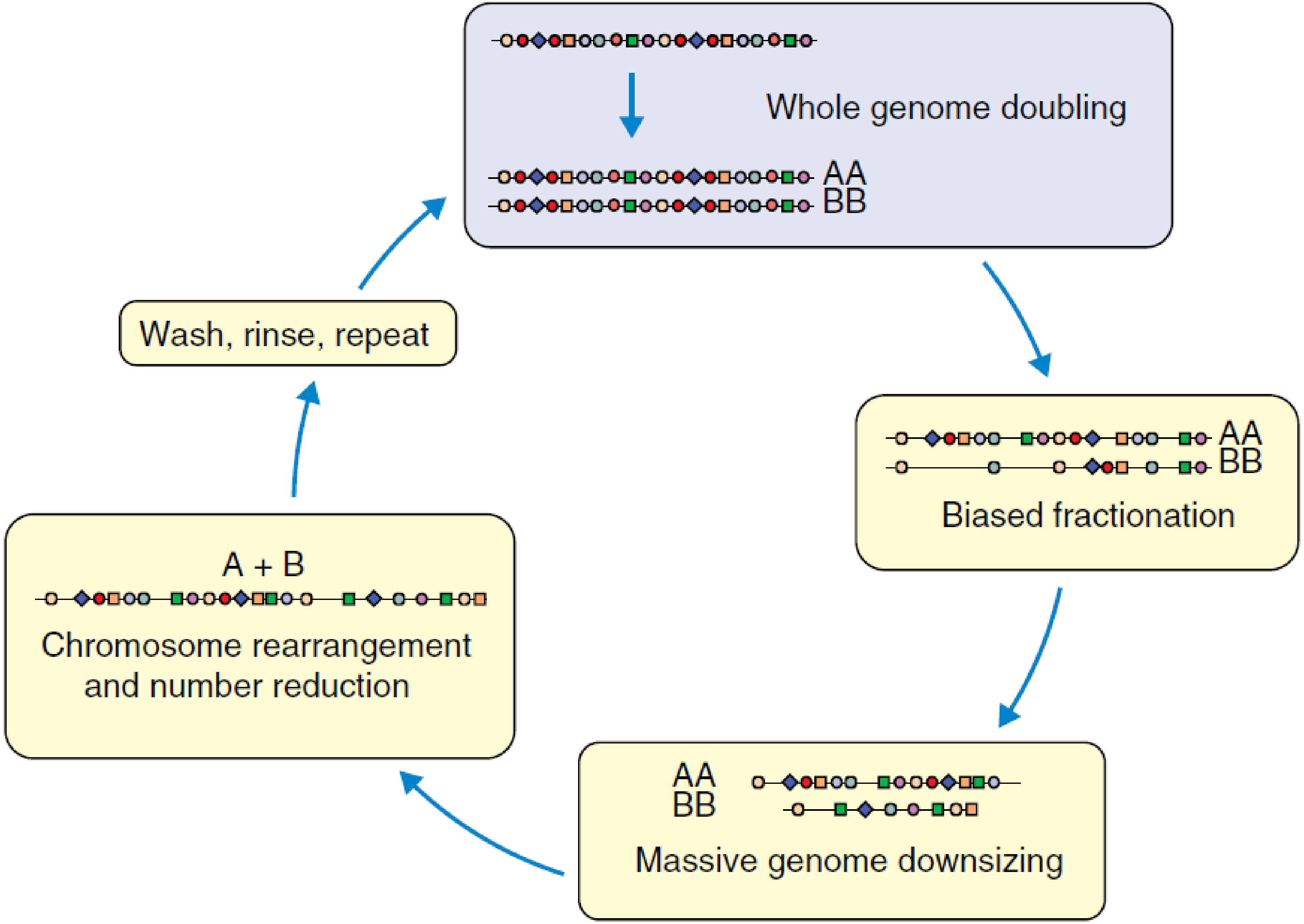
genome reshuffling

descending dysploidy (chromosome no. reduction)

genome downsizing and/or upsizing

diversification / species radiation

Whole-genome duplication and diploidization



Whole genome doubling

Biased fractionation

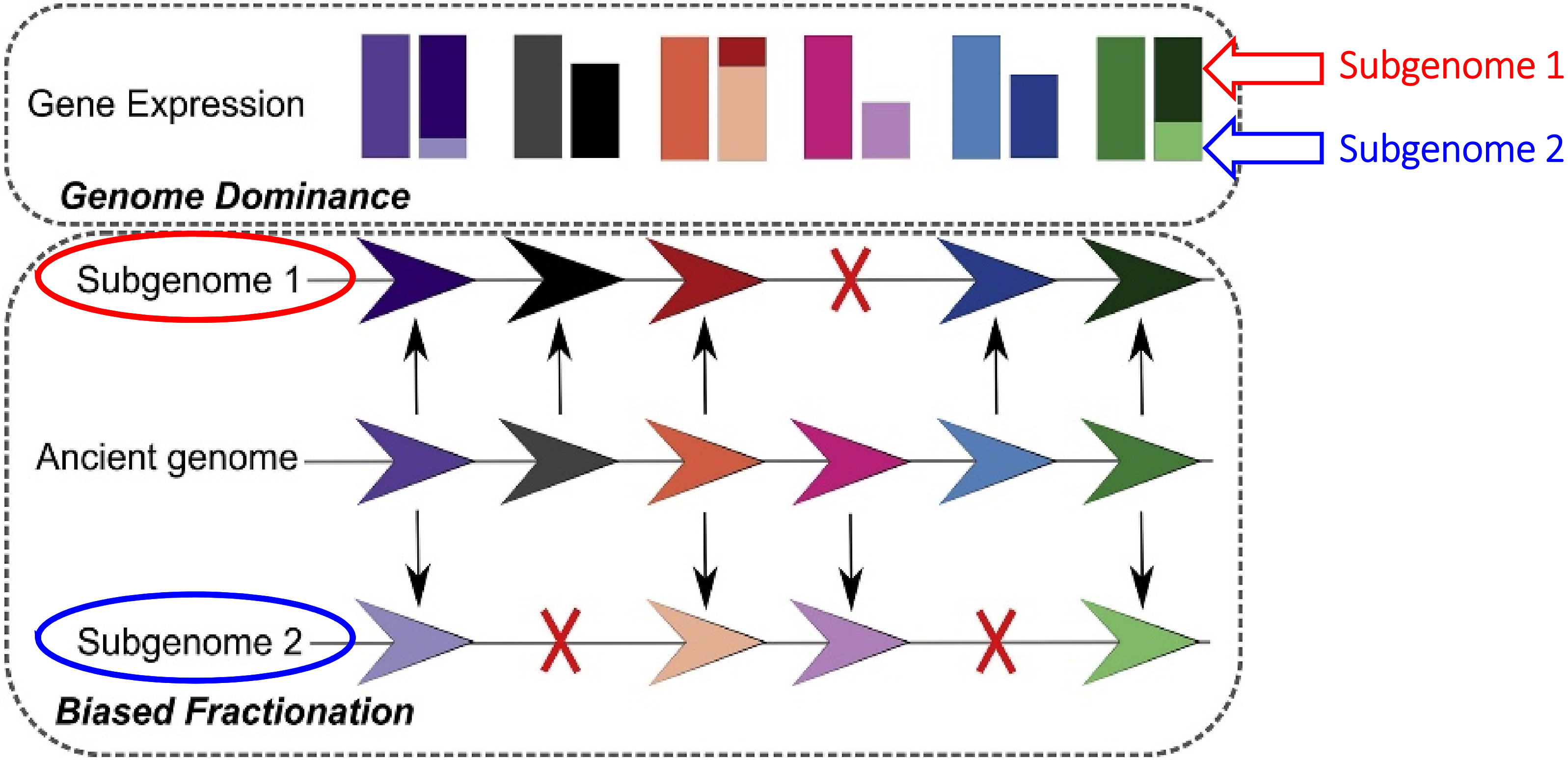
Massive genome downsizing

Chromosome rearrangement and number reduction

Wash, rinse, repeat

Wendel et al. (2016) Genome Biol

Genome diploidization: biased fractionation and (sub)genome dominance



Liang and Schnable (2018) Mol Plant

Biased (sub)genome fractionation and dominance can be explained by the mode of polyploidization

Garsmeur et al. (2013) Mol Biol Evol

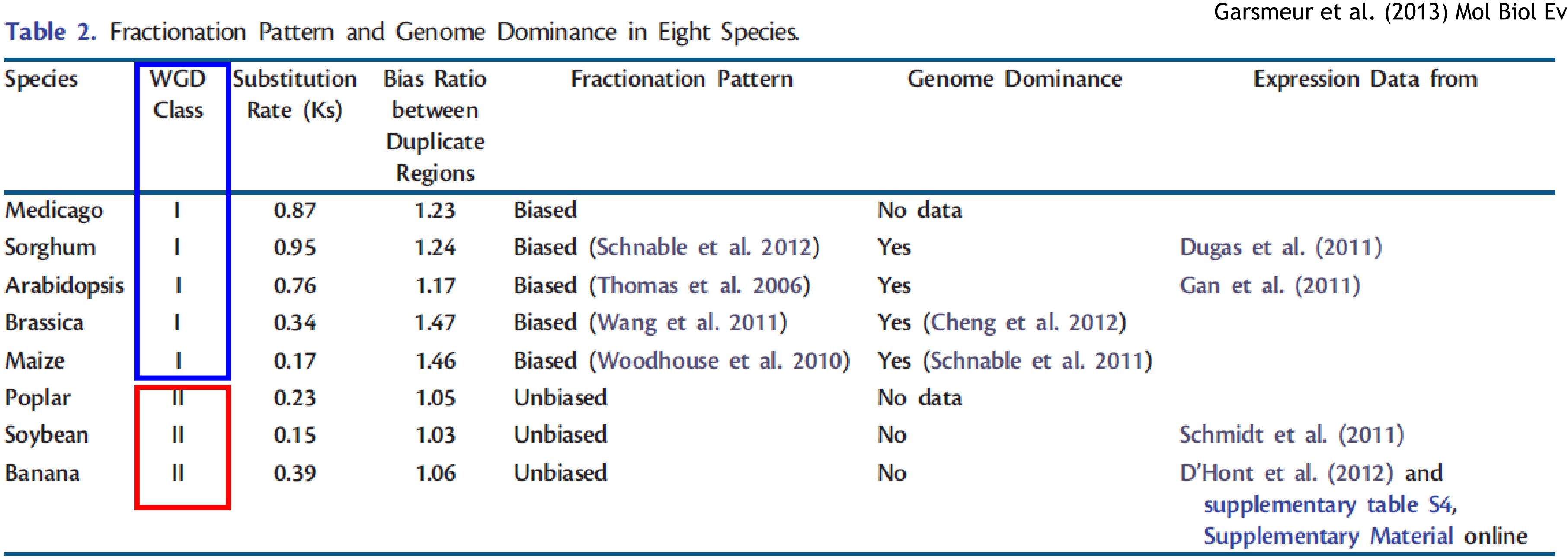
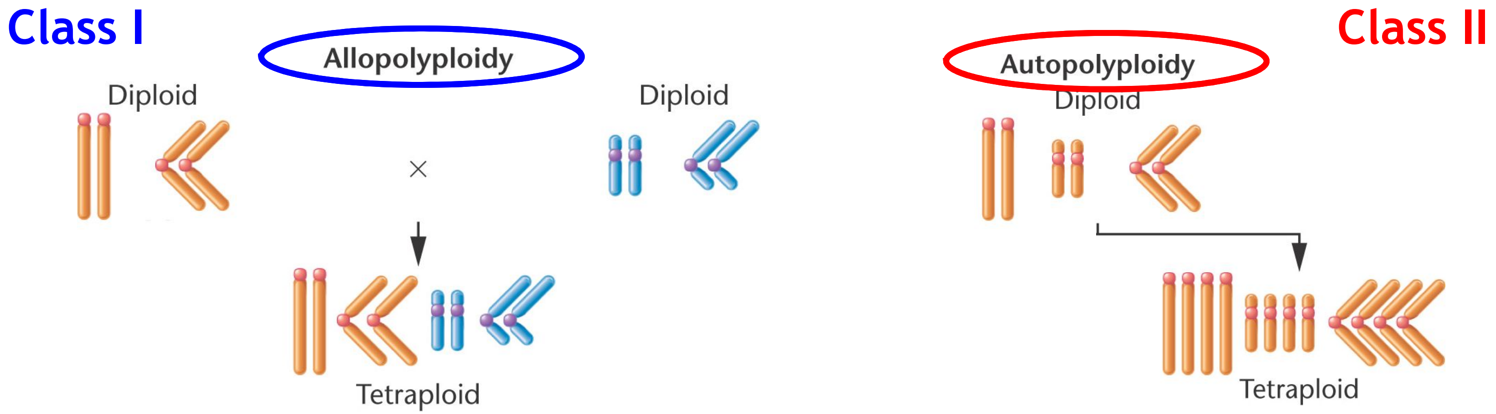


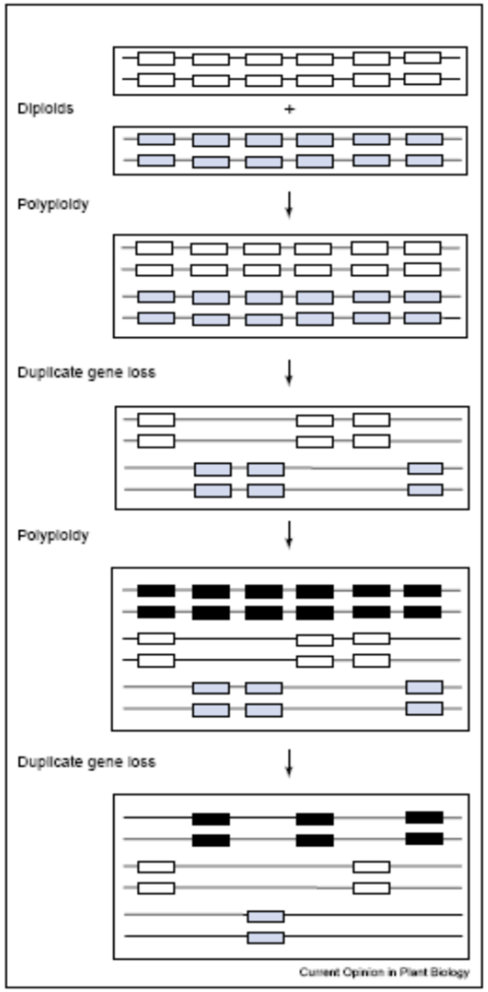
Table 2. Fractionation Pattern and Genome Dominance in Eight Species.

Species	WGD Class	Substitution Rate (Ks)	Bias Ratio between Duplicate Regions	Fractionation Pattern	Genome Dominance	Expression Data from
Medicago	I	0.87	1.23	Biased	No data	
Sorghum	I	0.95	1.24	Biased (Schnable et al. 2012)	Yes	Dugas et al. (2011)
Arabidopsis	I	0.76	1.17	Biased (Thomas et al. 2006)	Yes	Gan et al. (2011)
Brassica	I	0.34	1.47	Biased (Wang et al. 2011)	Yes (Cheng et al. 2012)	
Maize	I	0.17	1.46	Biased (Woodhouse et al. 2010)	Yes (Schnable et al. 2011)	
Poplar	II	0.23	1.05	Unbiased	No data	
Soybean	II	0.15	1.03	Unbiased	No	Schmidt et al. (2011)
Banana	II	0.39	1.06	Unbiased	No	D'Hont et al. (2012) and supplementary table S4, Supplementary Material online

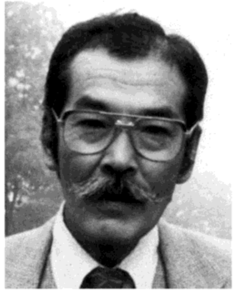


The fate of duplicated genes

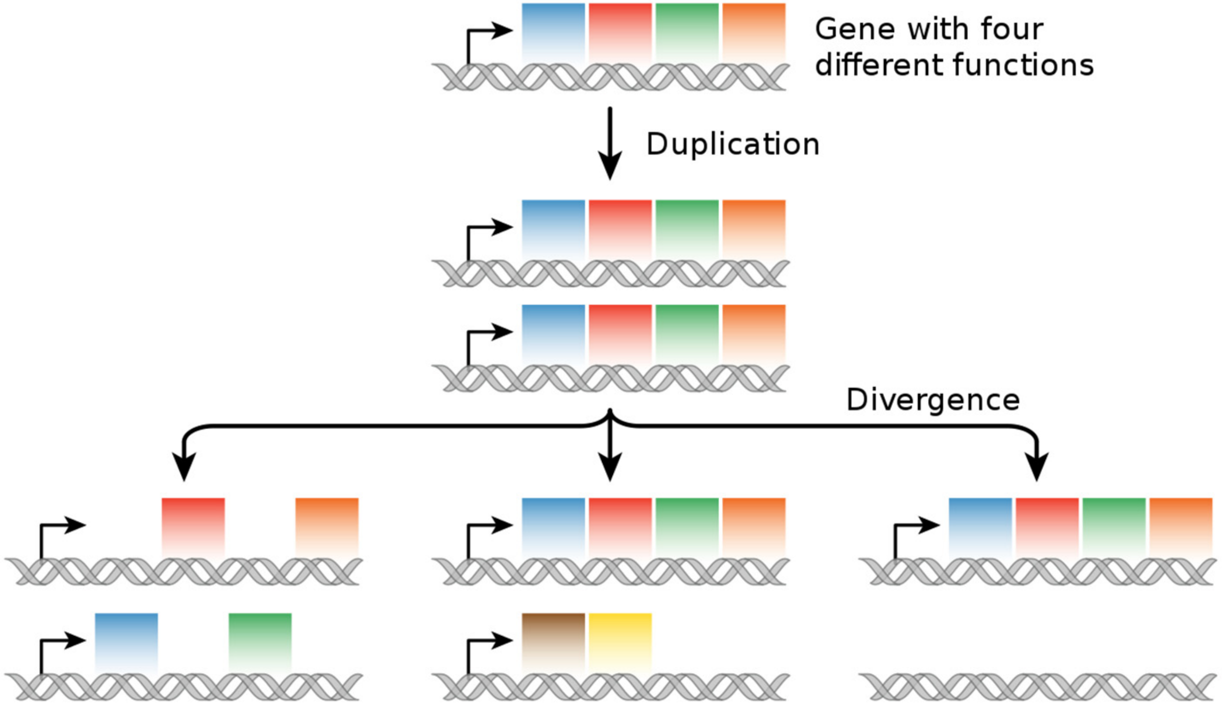
Genome evolution through cyclic
WGD and diploidization



Adams and Wendel (2005)



Susumu Ohno



Gene with four different functions

Duplication

Divergence

Subfunctionalization

Neofunctionalization

Degeneration/Gene loss

Genome size decrease (downsizing)

- recombination
- chromosome rearrangements

Genome size decrease (downsizing)

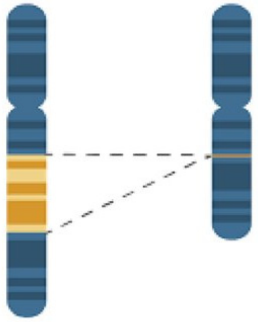
Recombinational deletions after double-strand breaks (DSBs) - DSB repair

- unequal homologous recombination including unequal crossing-over

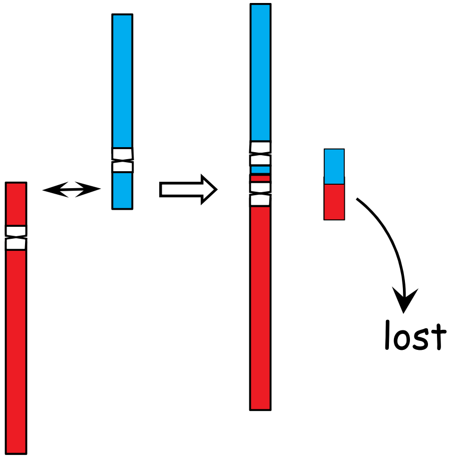
- illegitimate recombination (non-homologous end joining, NHEJ)

Chromosome rearrangements (…in principle again DSBs and recombination)

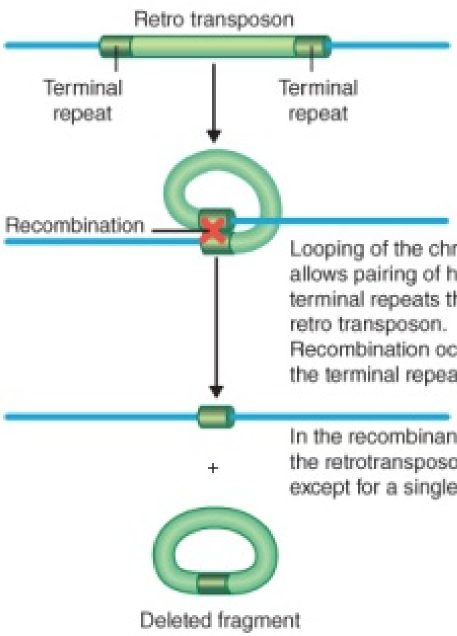
Large-scale deletion



Robertsonian translocation

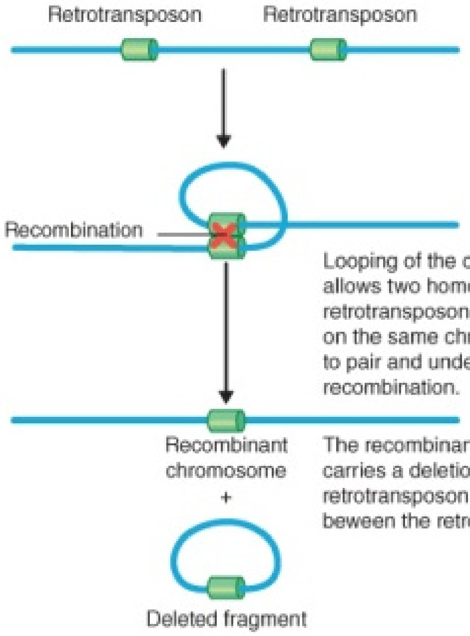


Genome size decrease by unequal homologous recombination between two LTRs or between two LTR-retrotransposons



Looping of the chromosome allows pairing of homologous terminal repeats that flank the retro transposon. Recombination occurs between the terminal repeats.

In the recombinant chromosome the retrotransposon is deleted except for a single terminal repeat.

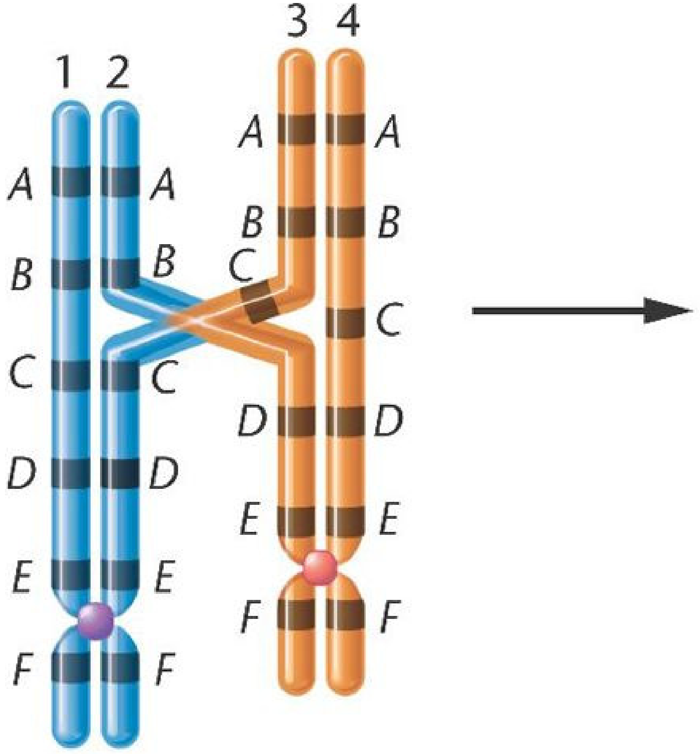


Looping of the chromosome allows two homologous retrotransposons located on the same chromosome to pair and undergo recombination.

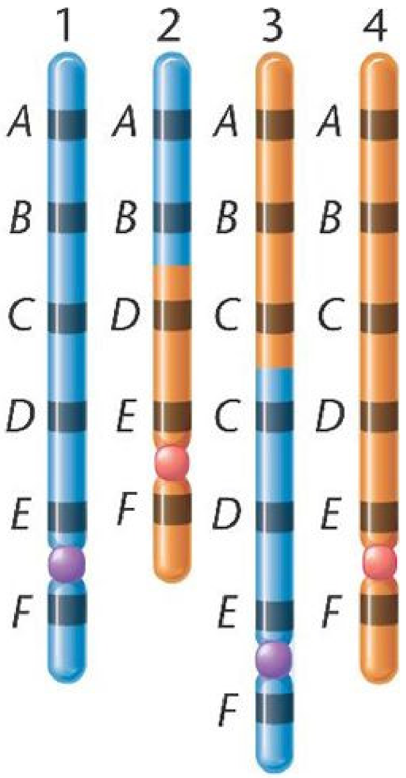
The recombinant chromosome carries a deletion of one retrotransposon and the region beween the retrotransposons.

~70% of retrotransposon sequences in the *A. thaliana* genome are no longer autonomous: solo LTRs = probably the consequence of unequal homologous recombination = inactive, truncated elements cannot contribute to genome expansion

Deletion through unequal crossing-over



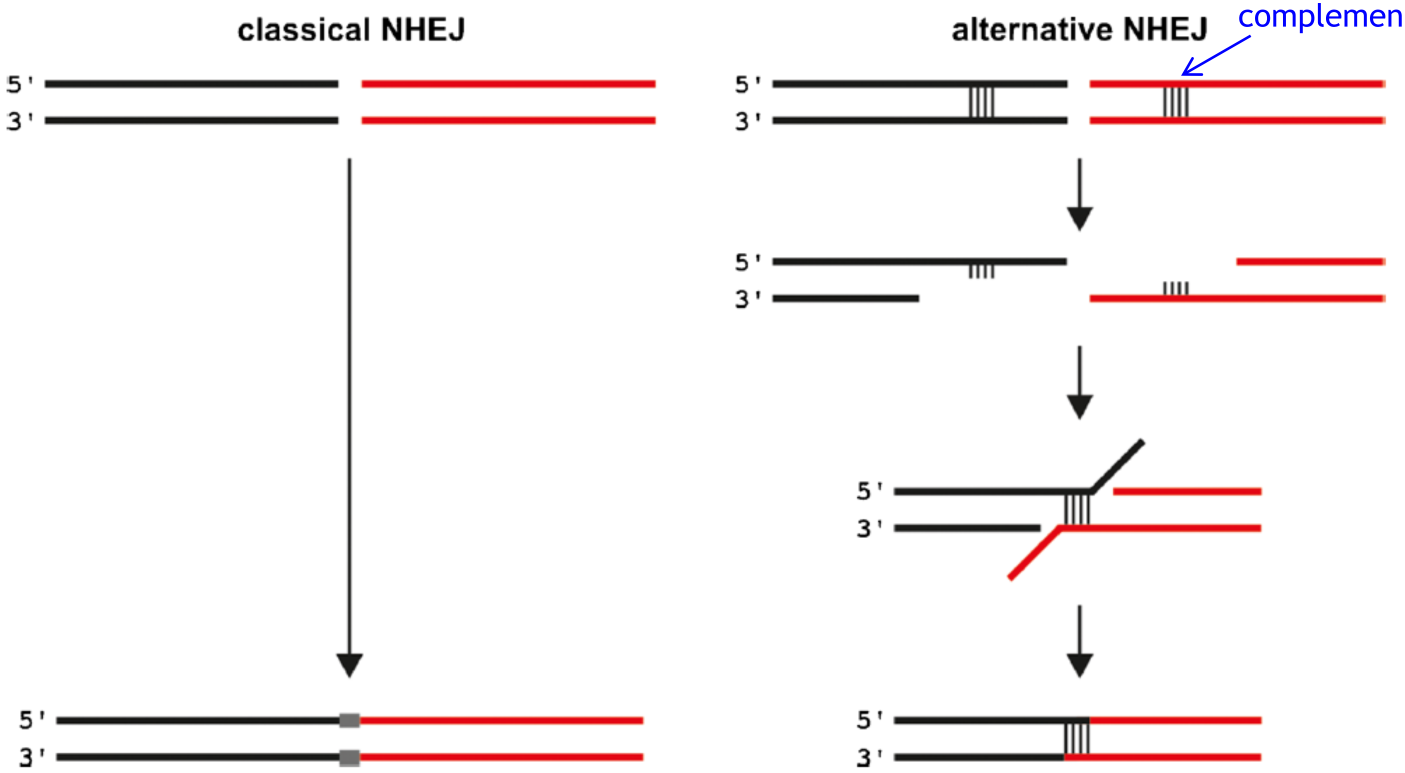
This tetrad is mispaired at meiotic synapsis.



The result, after crossing over, is two unequal chromosomes: one with a duplication (3) and one with a deletion (2).

Two main pathways of non-homologous end joining (NHEJ)

microhomology-mediated end joining (MMEJ)



complementary nucleotides

DNA lost

(but some DNA can be inserted - filler DNA)

NHEJ in plant somatic cells

- NHEJ seems to be the main mode of DSB repair in higher eukaryotes
- NHEJ might lead, in some cases, to genomic changes (deletions, insertions or various kinds of genomic rearrangements)
- genomic alterations in meristematic cells can be transferred to the offspring
- **alternative NHEJ can mediate genome size loss**

⇩

1C = 157 Mb

1C = 4.5 Gb

Arabidopsis vs. tobacco (genome size larger in tobacco)

- tobacco: almost every second deletion event is accompanied by the insertion of filler sequence

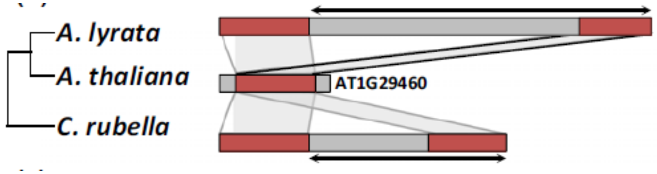
- Arabidopsis: no insertions

- overall length of the deletions is about one-third shorter in tobacco than in Arabidopsis

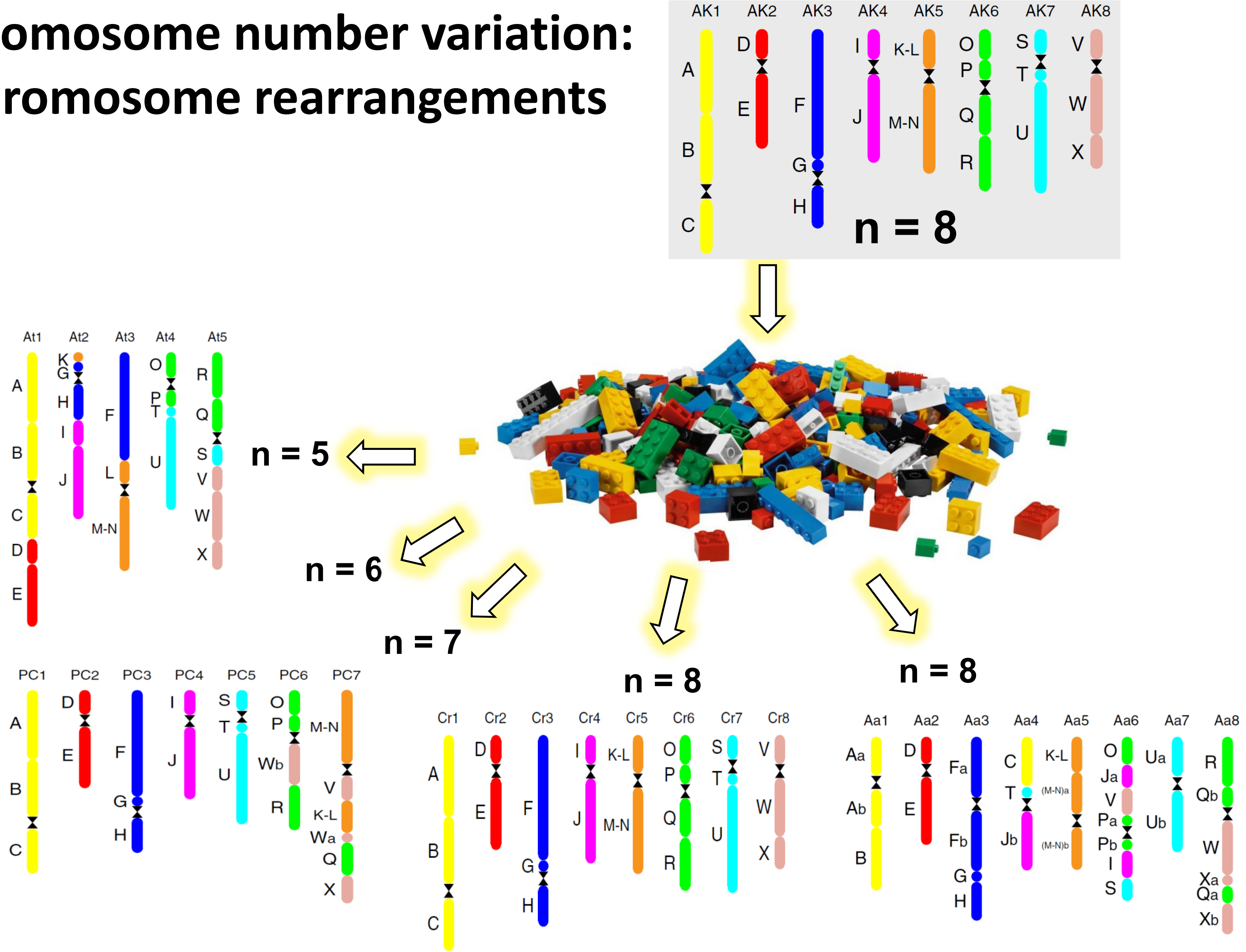
>>> inverse correlation between genome size and the medium length of deletions

>>> ??? species-specific differences in DSB repair pathways can contribute to the evolution of eukaryotic genome size ???

- *A. thaliana* (157 Mb) has lost **6**× more introns than *Arabidopsis lyrata* (210 Mb) since the divergence of the two species but gained very few introns

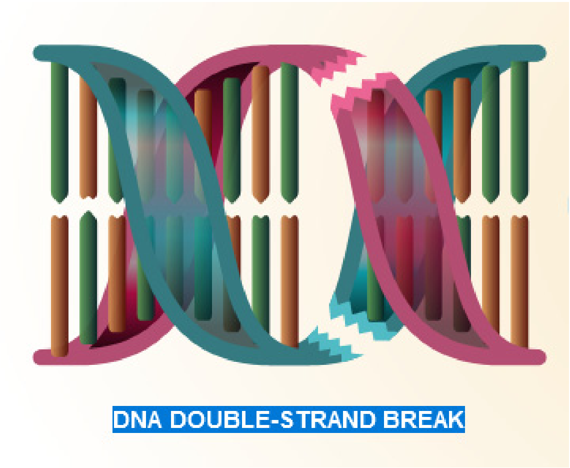


Chromosome number variation: chromosome rearrangements



Chromosome rearrangements results from double-strand breaks and their miss-repair

DSB



↓

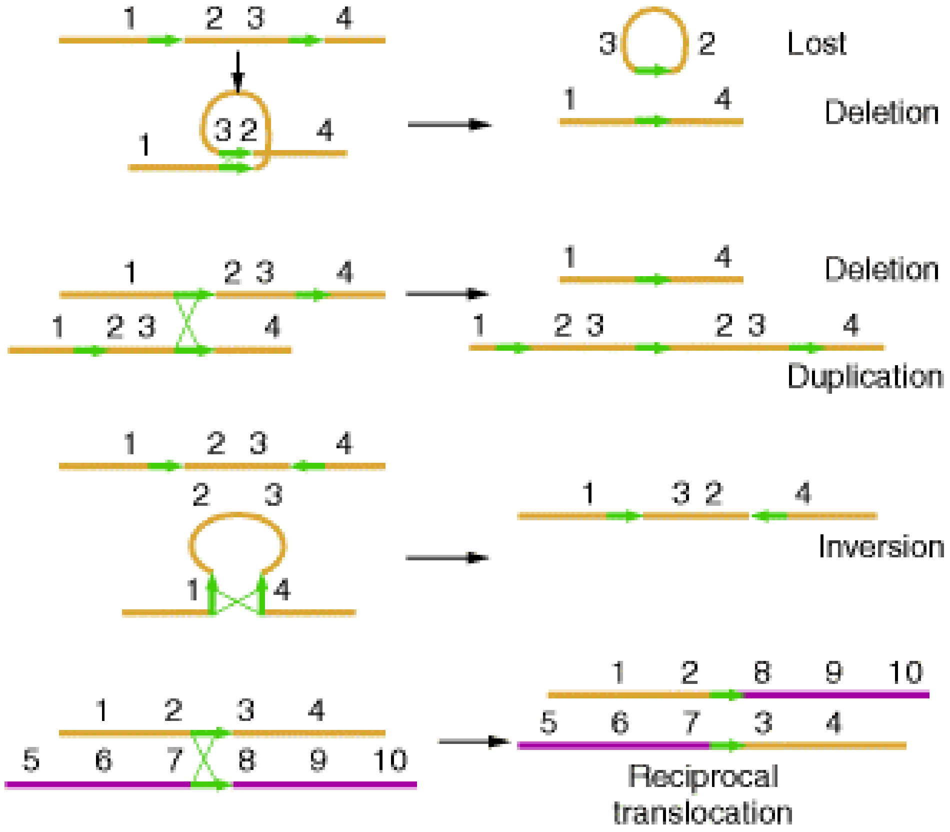
miss-repair

↓

chromosome rearrangement

Chromosome rearrangements – the role of repeats

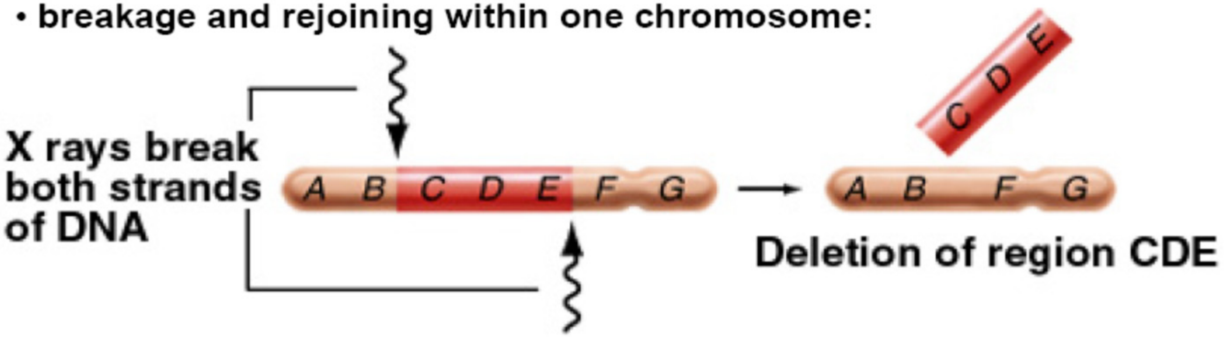
In organisms with repetitive DNA, homologous repetitive segments **within one chromosome** or **on different chromosomes** can act as sites of DSBs and their miss-repair, i.e. non-allelic homologous recombination.

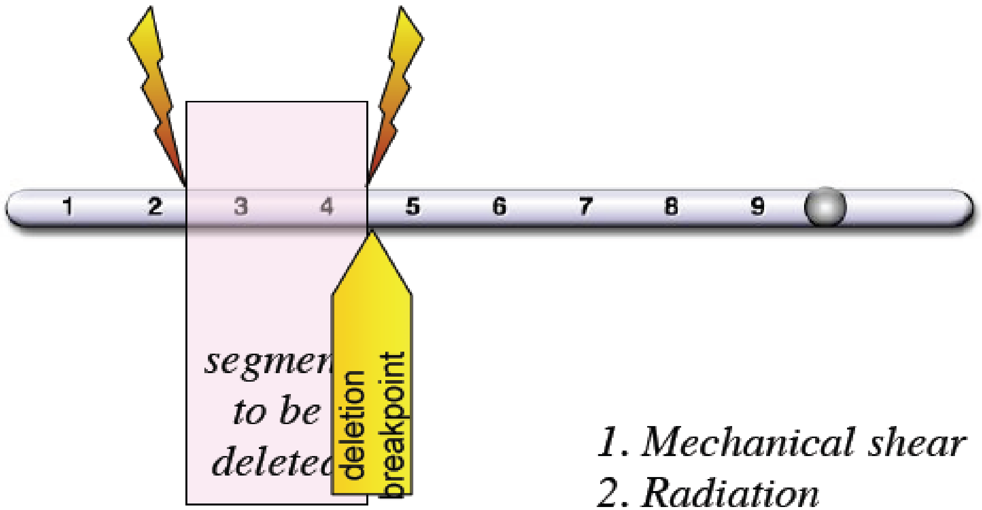


Deletion formation by breakage and rejoining

= deficiencies = losses of chromosome segments

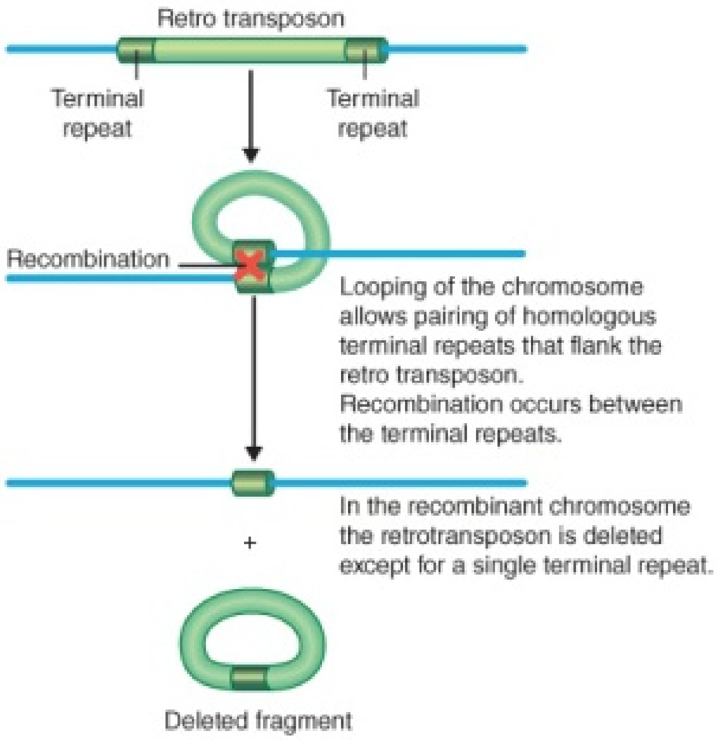
- can occur terminally or internally, e. g. caused by…
  - breakage and rejoining within one chromosome:





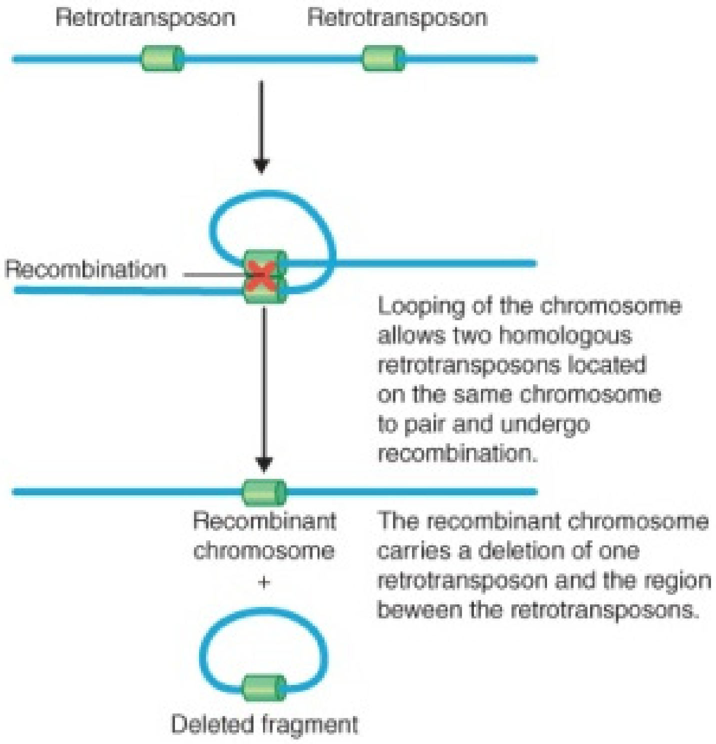
1. Mechanical shear
2. Radiation
3. Transposable elements

Deletion formation by intra-chromosomal (unequal) recombination



Looping of the chromosome allows pairing of homologous terminal repeats that flank the retro transposon. Recombination occurs between the terminal repeats.

In the recombinant chromosome the retrotransposon is deleted except for a single terminal repeat.

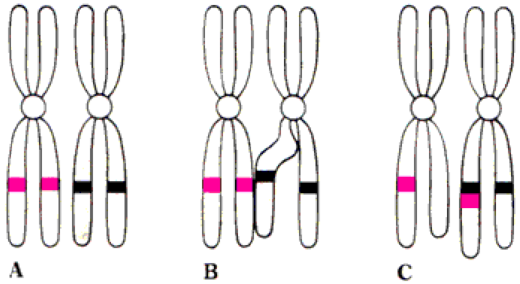


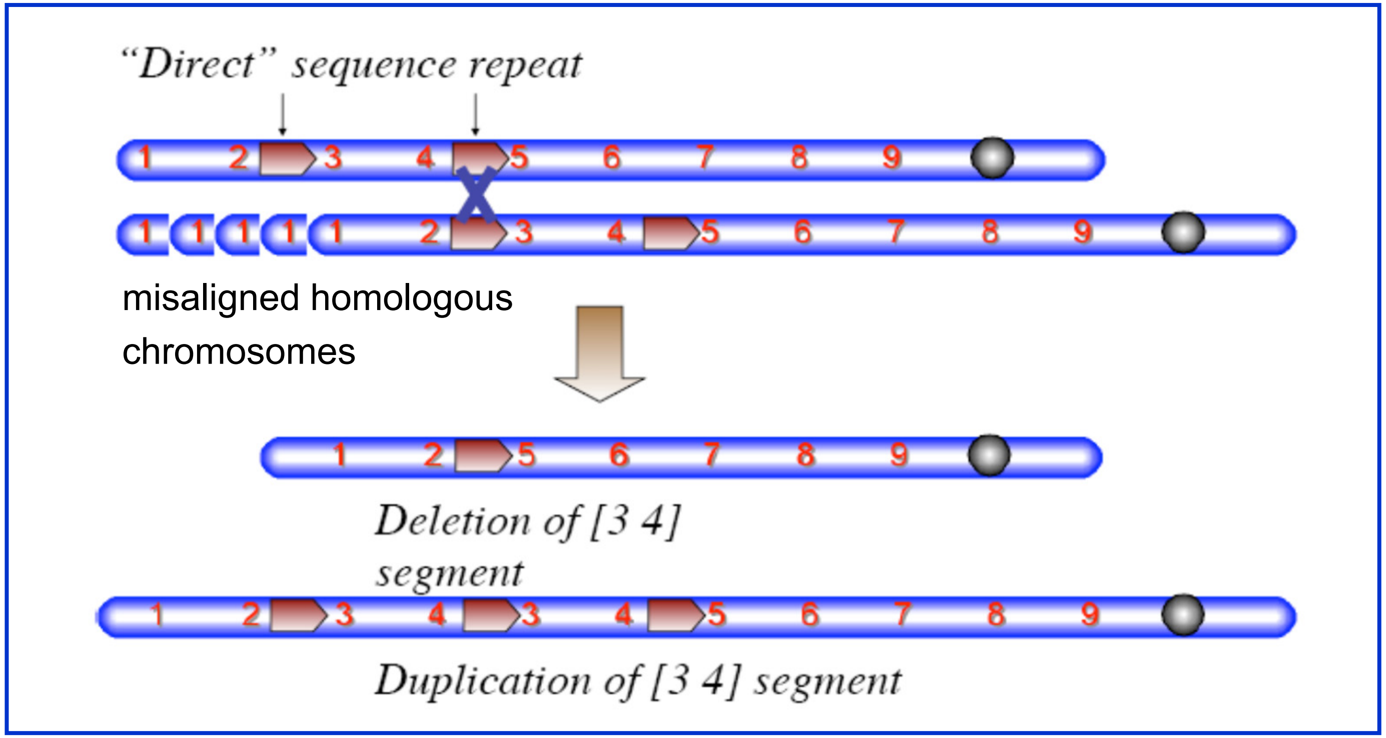
Looping of the chromosome allows two homologous retrotransposons located on the same chromosome to pair and undergo recombination.

The recombinant chromosome carries a deletion of one retrotransposon and the region beween the retrotransposons.

Deletion (and duplication) formation by unequal cross-over

Sometimes during meiosis two chromatids from homologous chromosomes **(A)** are misaligned during a cross-over event **(B)** as a result, one chromatid gained a duplicated region and the another lost a deleted region **(C)**. The duplication as well as the deletion are inherited by resulting gametes.





"Direct" sequence repeat

misaligned homologous chromosomes

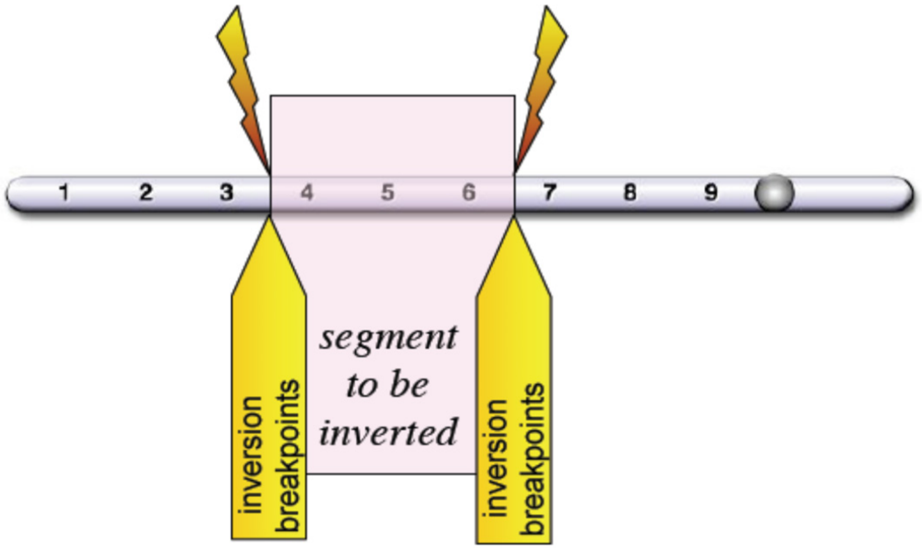
Deletion of [3 4] segment

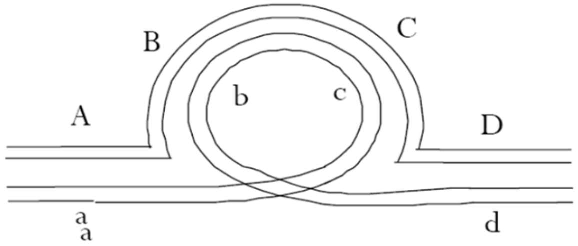
Duplication of [3 4] segment

Inversions

Inversions as balanced rearrangements are generally viable and show no particular abnormalities at the phenotypic level. Many inversions can be made homozygous.

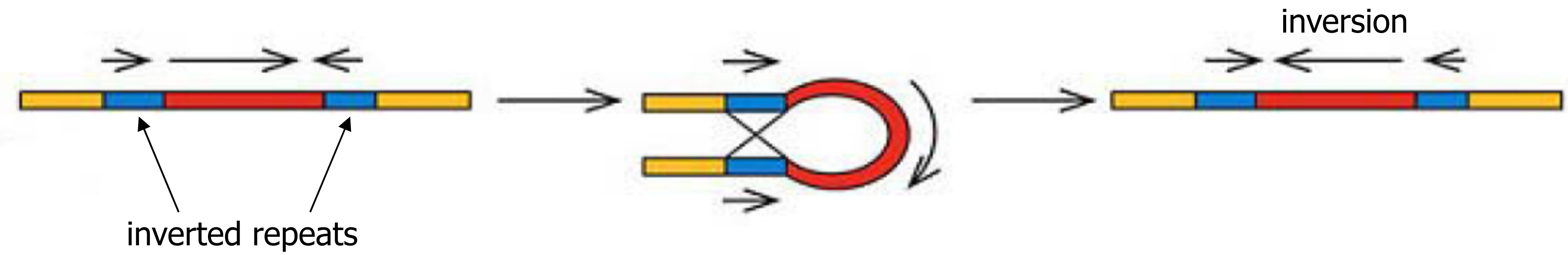
Inversion heterozygote - cells that contain one normal haploid chromosome set plus one set carrying the inversion. Microscopic observation of meioses in inversion heterozygotes reveals an **inversion loop**.





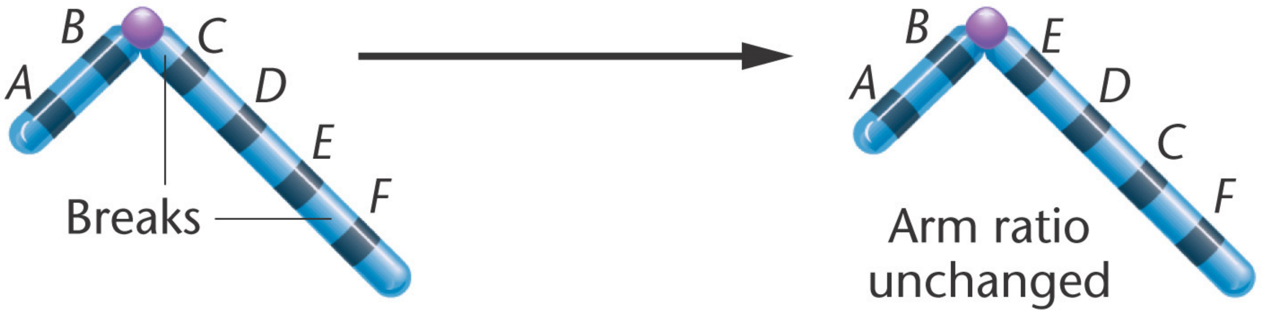
meiotic inversion loop

Inversion formation by intra-chromosomal recombination

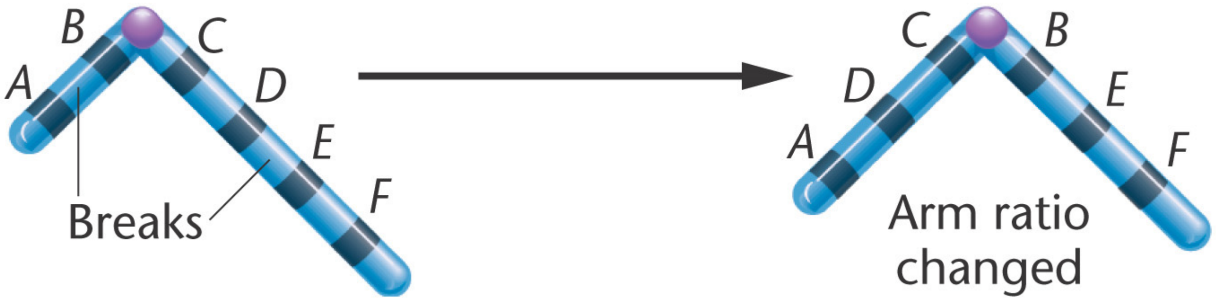


Two types of inversions

Paracentric inversion



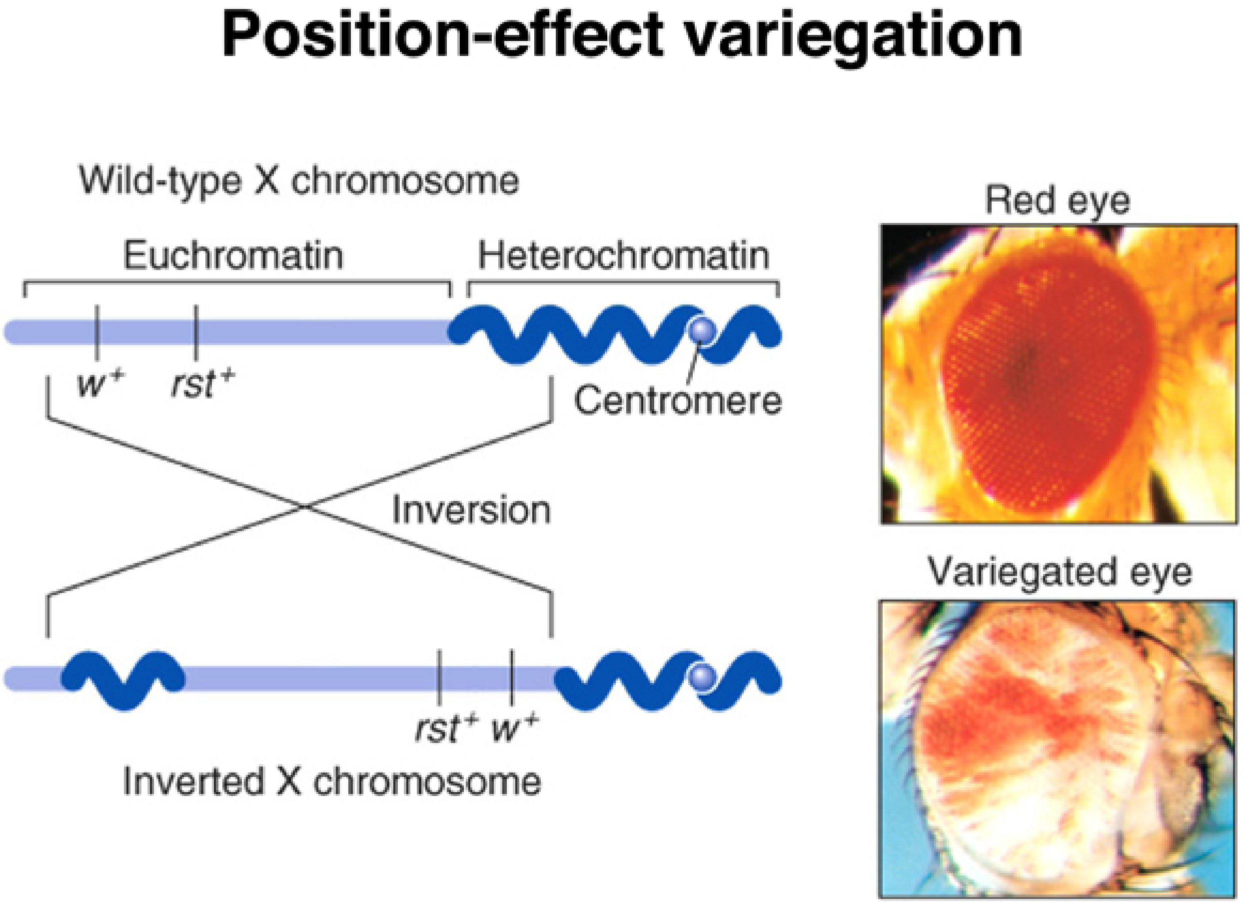
Pericentric inversion



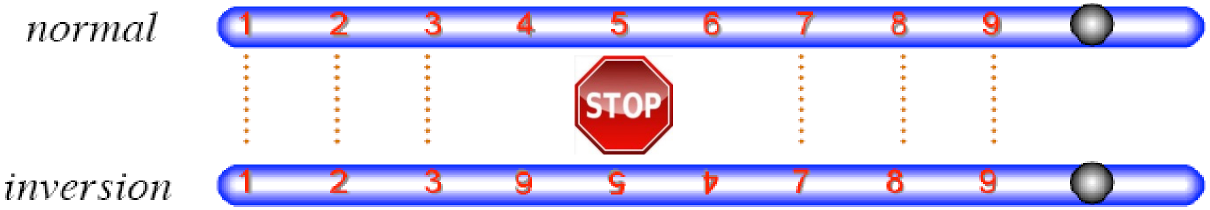
mechanism of inversion formation: breakage and rejoining

Inversions and recombination: evolutionary significance

Can be “adaptive” when it stabilizes/disrupt a superior combination of alleles on a chromosome (examples seen in *Drosophila)*



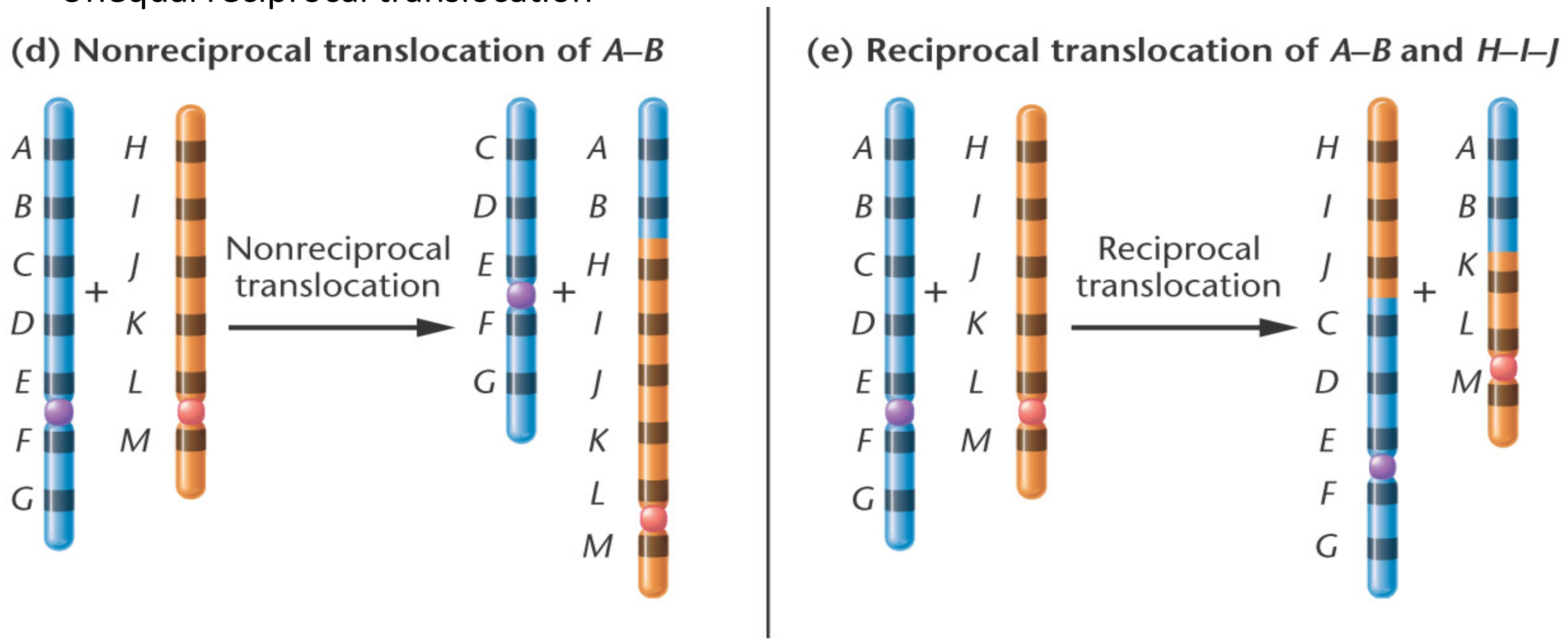
Inversions may suppress recombination



Chromosome rearrangements (typically inversions) may reduce gene flow by suppressing recombination. Inversions allow genes located in these regions to differentiate, in contrast to genes in freely recombining collinear regions.

Reciprocal translocations

Unequal reciprocal translocation



attachment of chromosome fragment to a non-homologous chromosome (leading to deletions and duplications in the progeny)

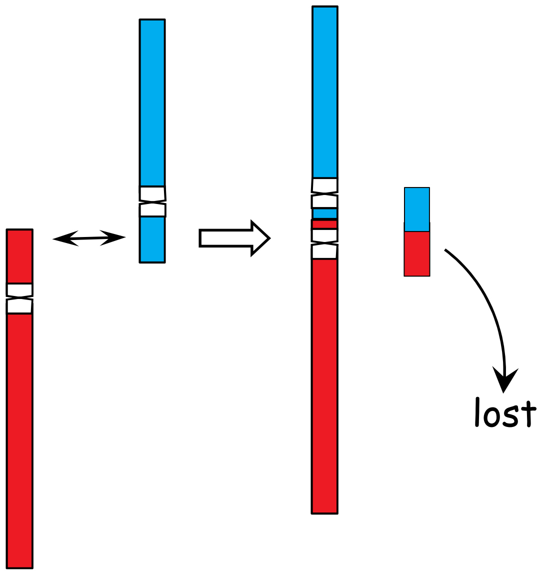
exchange of chromosome fragments between non-homologous chromosomes

Robertsonian translocations - ROBs (centric „fusions“)

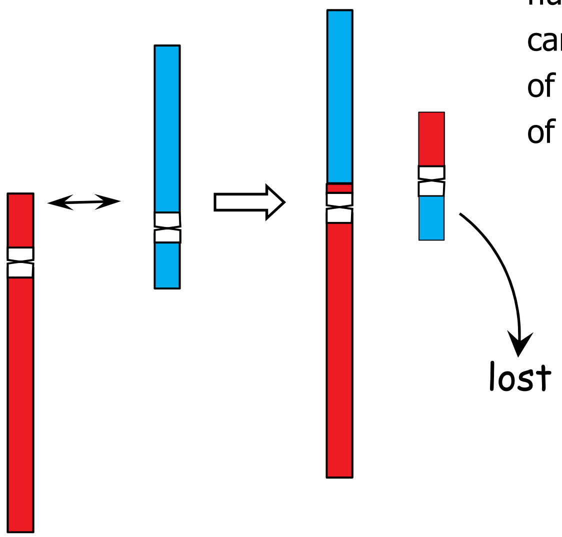
- type of a reciprocal translocation between two acrocentric/telocentric chromosomes
- also called whole-arm translocations or centric-fusion translocations
- named after the American insect geneticist W. R. B. Robertson, who first described a Robertsonian translocation in grasshoppers in 1916
- evolutionary significance >>> **chromosome number reduction** (from 2 acrocentric chromosomes one metacentric chromosome)

Dicentric ROB

(more frequent)

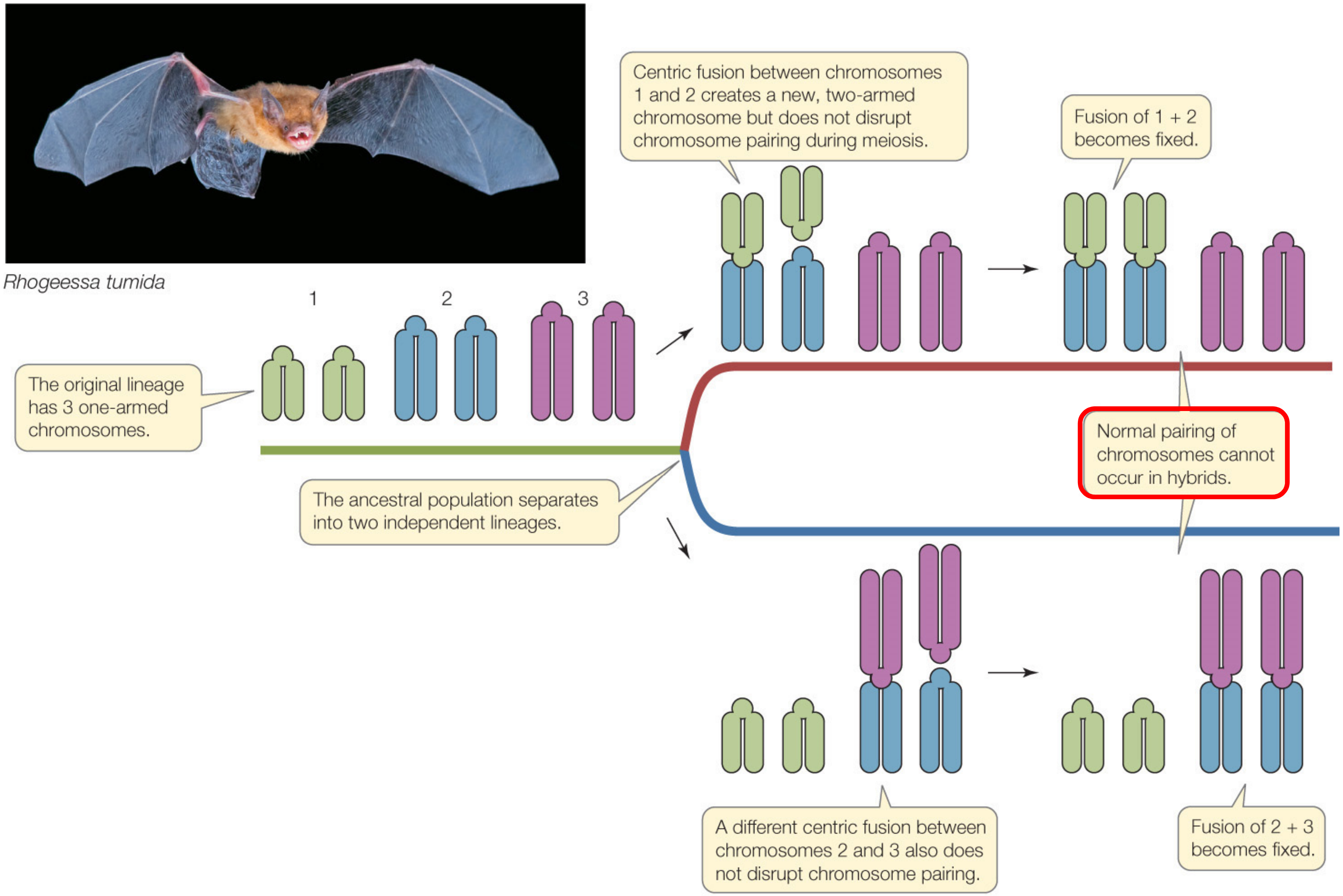


Monocentric ROB



Robertsonian translocations are the most common recurrent structural anomaly in humans, with about 1 in 1000 individuals carrying this rearrangement. The carriers of ROBs have 45 chromosomes instead of the normal 46.

Speciation by Robertsonia translocations („centric fusions“)



Rhogeessa tumida

The original lineage has 3 one-armed chromosomes.

The ancestral population separates into two independent lineages.

Centric fusion between chromosomes 1 and 2 creates a new, two-armed chromosome but does not disrupt chromosome pairing during meiosis.

Fusion of 1 + 2 becomes fixed.

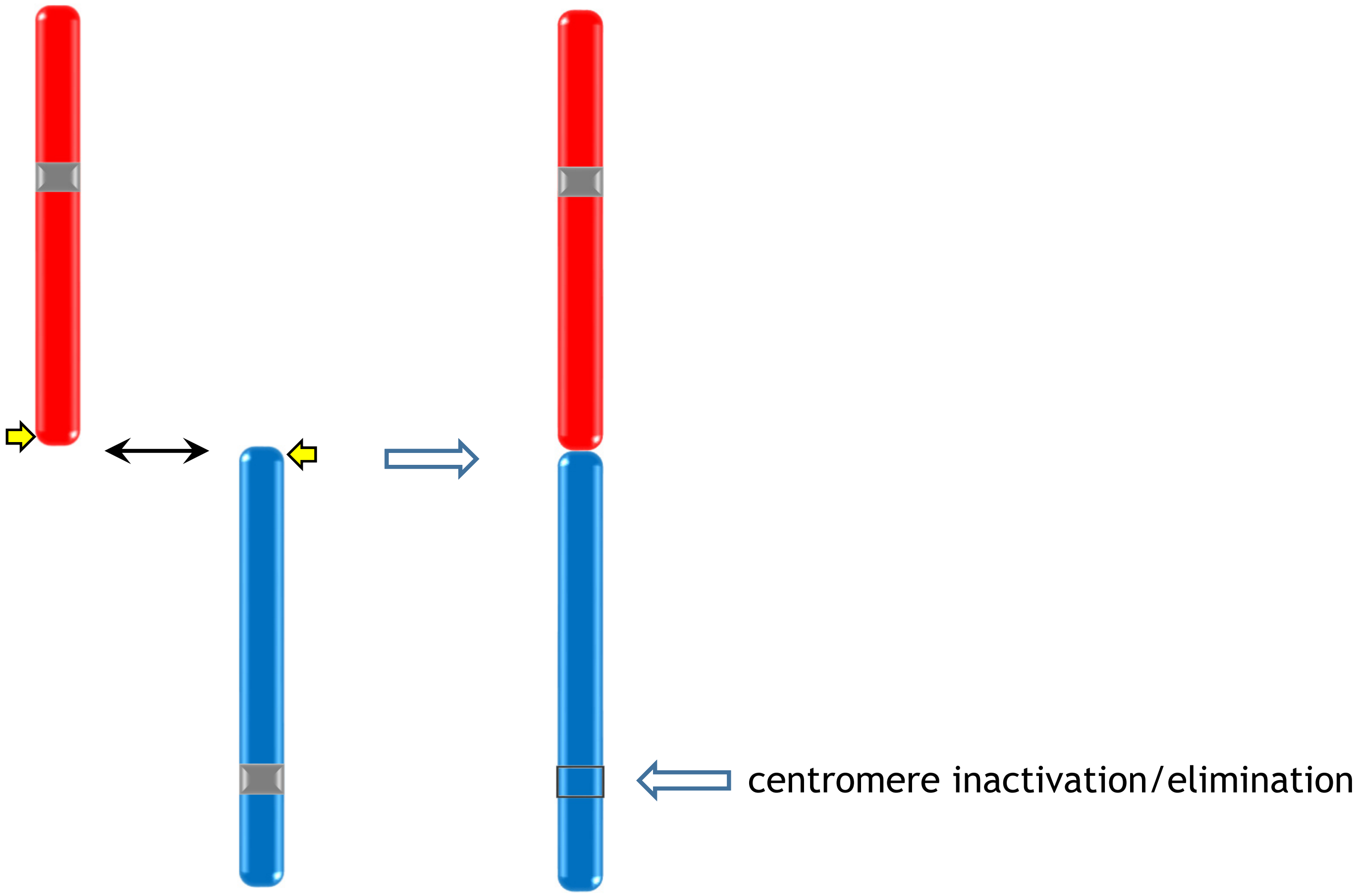
Normal pairing of chromosomes cannot occur in hybrids.

A different centric fusion between chromosomes 2 and 3 also does not disrupt chromosome pairing.

Fusion of 2 + 3 becomes fixed.

17.4: © Barry Mansell/Naturepl.com.

End-to-end chromosome translocations („chromosome fusions“)



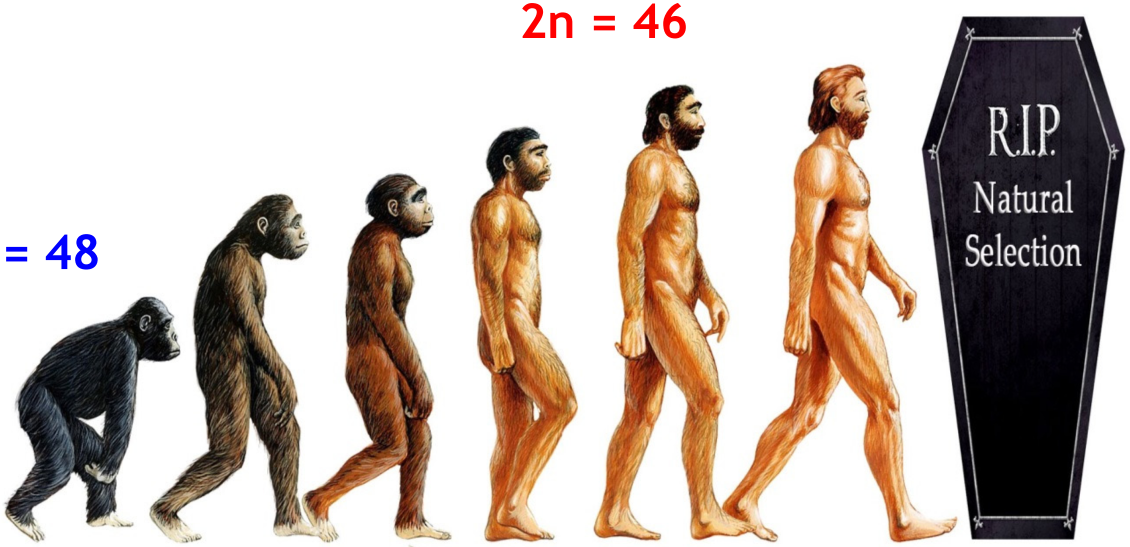
In principle unequal reciprocal translocation with breakpoints in (sub)telomeric regions.

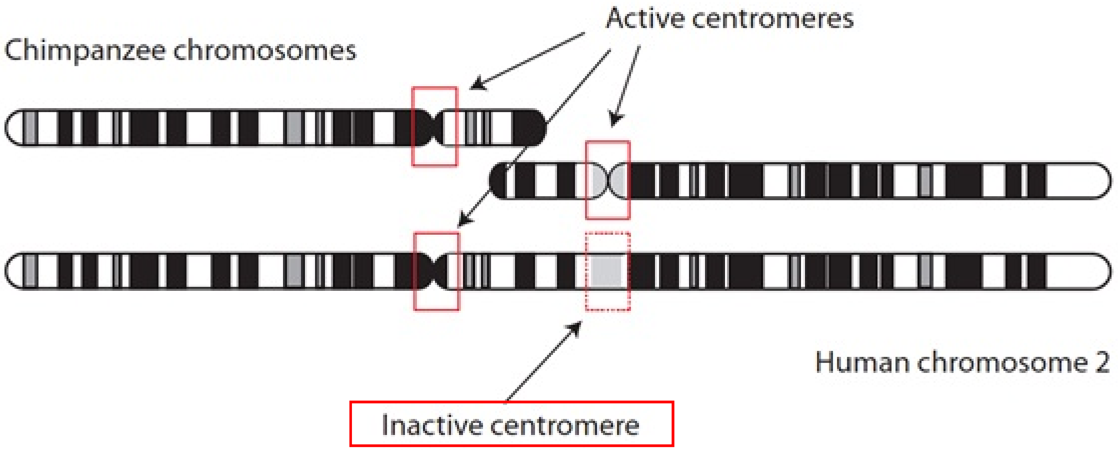
The second translocation product is minute and eliminated.

Chromosome „fusion“ – the origin of the human (dicentric) chromosome 2

2n = 46

2n = 48

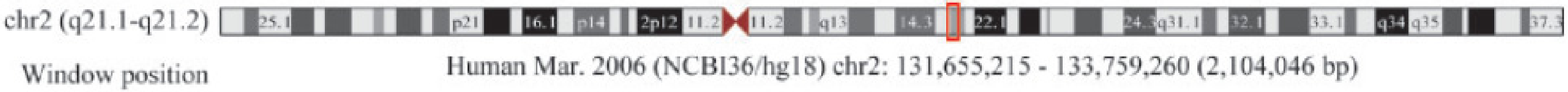


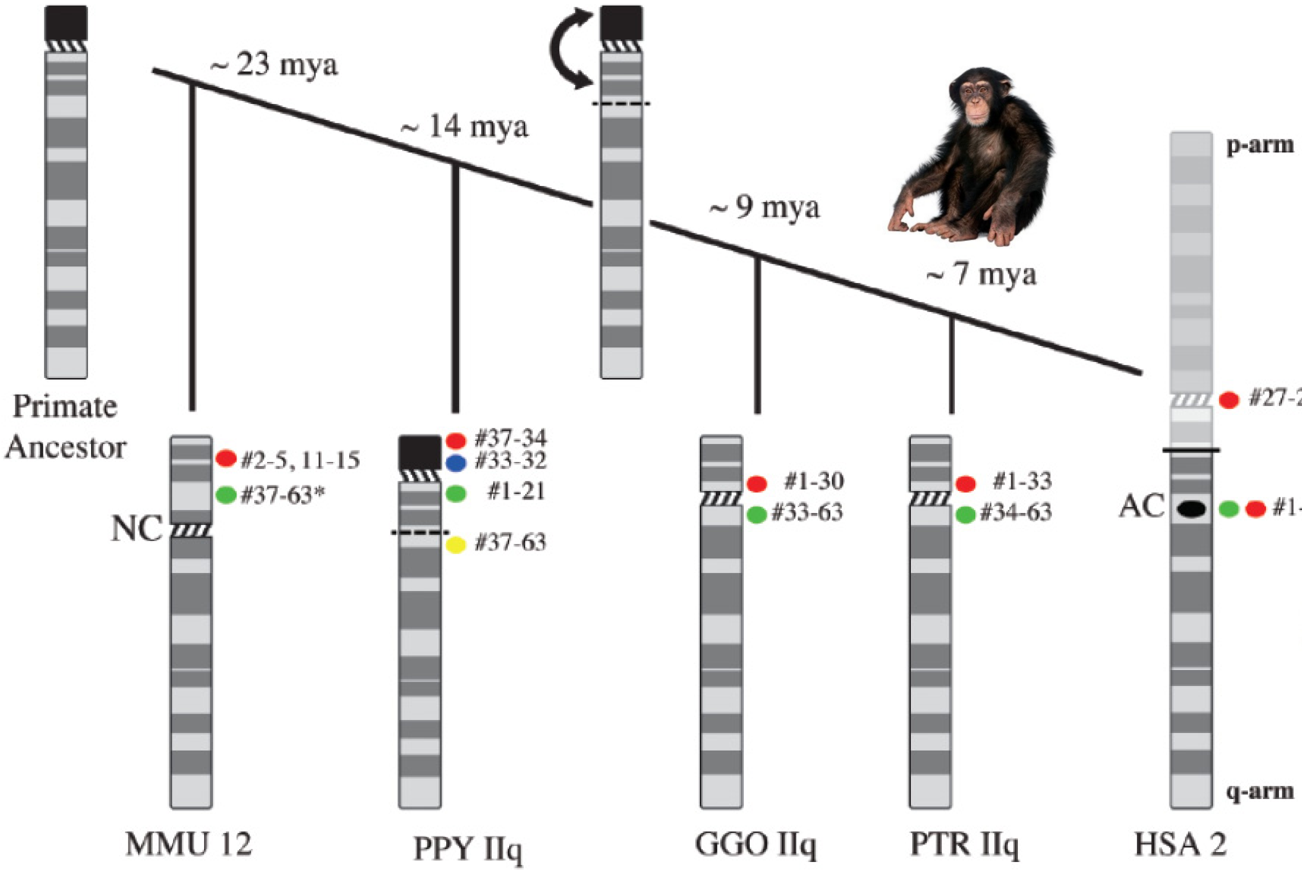


2n = 48

2n = 46

Chromosome „fusion“ – the origin of the human (dicentric) chromosome 2



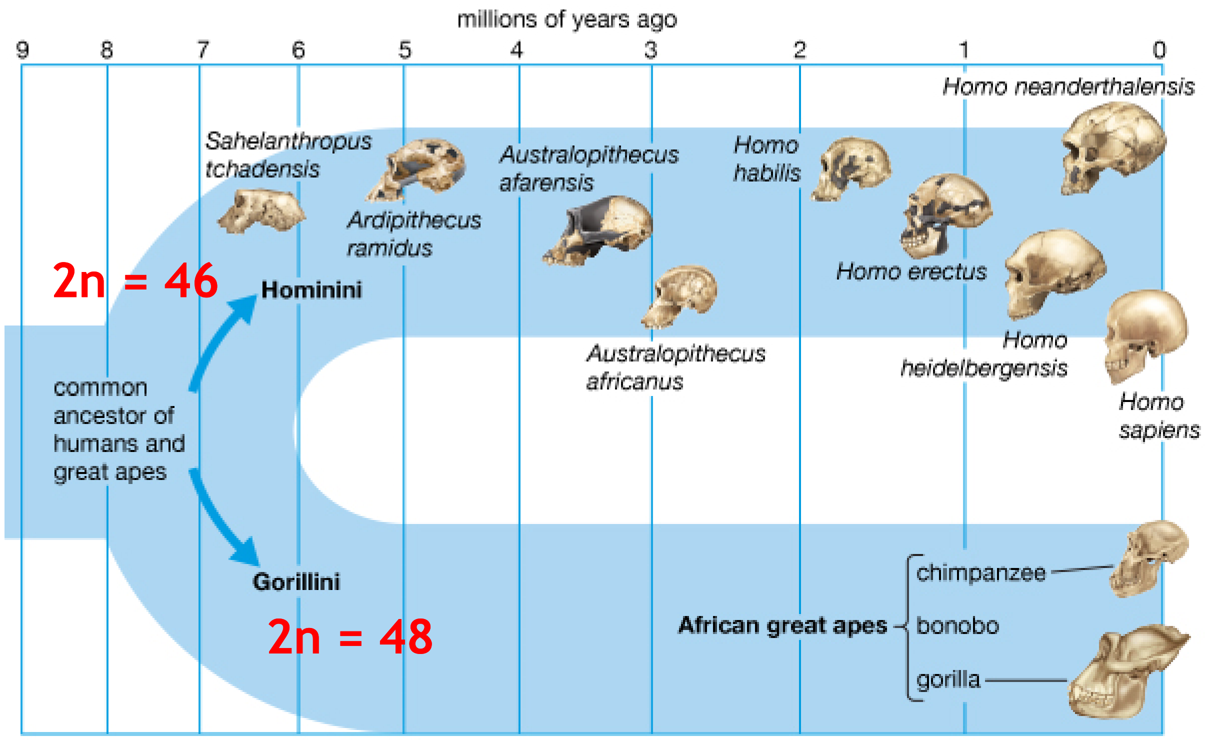


active centromere

inactive centromere

Chiatante et al. (2017), MBE

Did the origin of „fusion“ chromosome 2 contributed to reproductive isolation of hominid species from great apes?



- different no. of chromosomes → reproductive isolation
- loss of gene(s) → adaptive advantage
- gene linkage? changed regulation of gene expression?

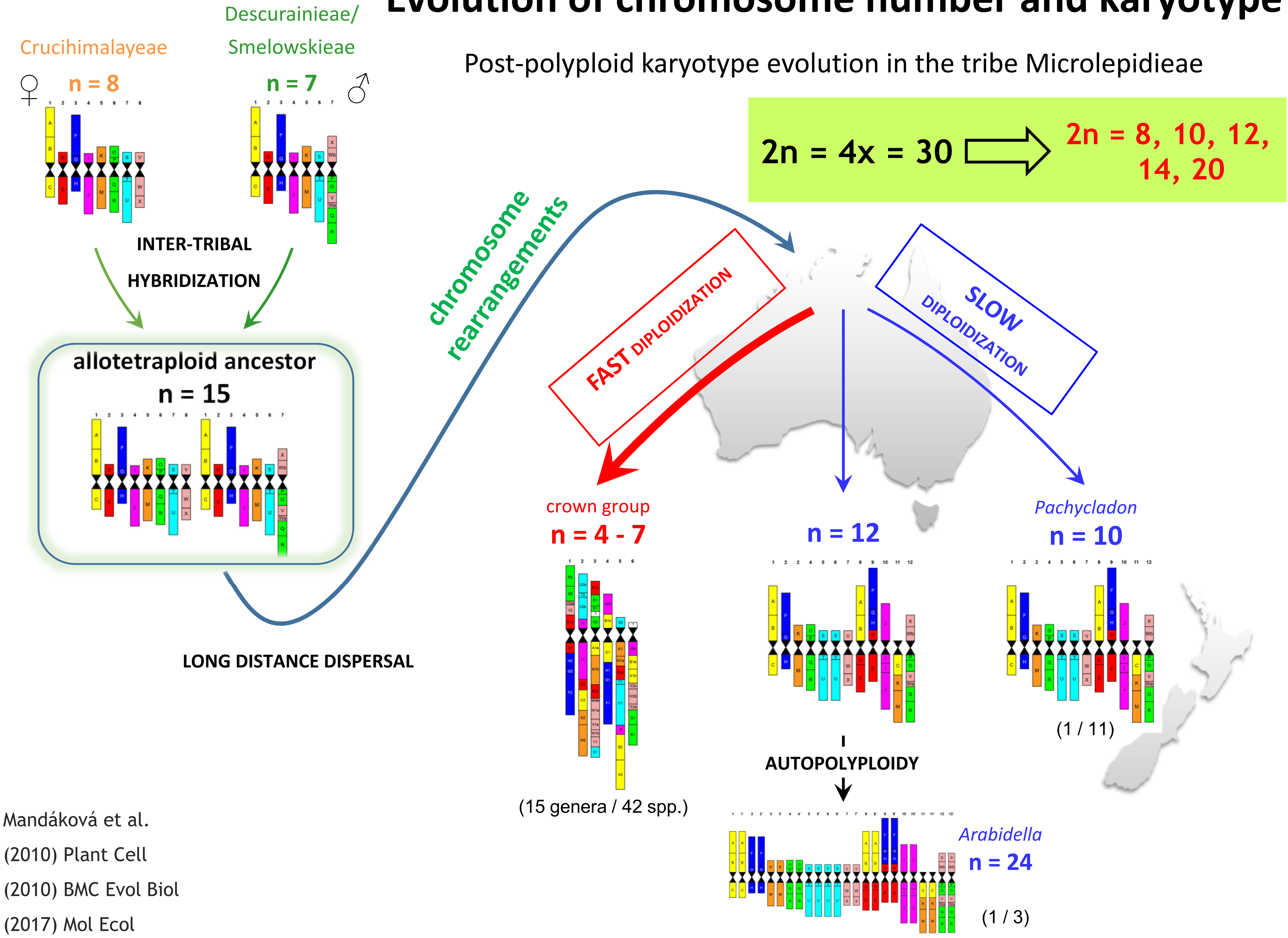
Evolution of chromosome number and karyotype

Allopolyploid origin and diploidization in the tribe Microlepidieae (Brassicaceae)

- Australia: 15 genera, 47 species
- New Zealand: *Pachycladon*, 11 species
- **chromosome number variation (from n = 4 to n = 24)**

Evolution of chromosome number and karyotype

Post-polyploid karyotype evolution in the tribe Microlepidieae



Crucihimalayeae

♀ n = 8

Descurainieae/
Smelowskieae

n = 7 ♂

INTER-TRIBAL

HYBRIDIZATION

allotetraploid ancestor
n = 15

LONG DISTANCE DISPERSAL

chromosome rearrangements

2n = 4x = 30 ⇨ 2n = 8, 10, 12, 14, 20

FAST DIPLOIDIZATION

SLOW DIPLOIDIZATION

crown group
n = 4 - 7

(15 genera / 42 spp.)

n = 12

AUTOPOLYPLOIDY

Arabidella
n = 24

(1 / 3)

Pachycladon
n = 10

(1 / 11)

Mandáková et al.
(2010) Plant Cell
(2010) BMC Evol Biol
(2017) Mol Ecol

martin.lysak@ceitec.muni.cz

one diploma & one doctoral student

www.plantcytogenomics.org